Systemic inequities in introductory physics courses: the impacts of learning assistants

Ben Van Dusen[1] and Jayson M. Nissen[1]

[1]*Department of Science Education, California State University Chico, Chico, CA, 95929, USA*

Creating equitable performance outcomes among students is a focus of many instructors and researchers. One focus of this effort is examining disparities in physics student performance across genders, which is a well-established problem. Another less common focus is disparities across racial and ethnic groups, which may have received less attention due to low representation rates making it difficult to identify gaps in their performance. In this investigation we examined associations between Learning Assistant (LA) supported courses and improved equity in student performance. We built Hierarchical Linear Models of student performance to investigate how performance differed by gender and by race/ethnicity and how LAs may have moderated those differences. Data for the analysis came from pre-post concept inventories in introductory mechanics courses collected through the Learning About STEM Student Outcomes (LASSO) platform. Our models show that gaps in performance across genders and races/ethnicities were similar in size and increased from pre to post instruction. LA-support is meaningfully and reliably associated with improvement in overall student performance but not with shifts in within-course performance gaps.

Keywords: Equity, Gender, Race, Ethnicity, Concept Inventory, LASSO, Learning Assistant

I. INTRODUCTION

Disparities in student performance in science classes have been recorded throughout the United States' educational systems [1]. The PER community has a significant number of publications [2] documenting the impact of course transformations on student performance overall, but has only begun to disaggregate these findings across student demographics [3]. The National Research Council (NRC) report examining the state of Discipline Based Education Research [2, pg. 136-137] states that while, "DBER clearly indicates that student-centered instructional strategies can positively influence students' learning...Most of the studies the committee reviewed were not designed to examine differences in terms of gender, ethnicity, socioeconomic status, or other student characteristics." The NRC identifies examining performance of students from underrepresented cultures as an important direction for future research.

In one of the few investigations of equity for both gender and race/ethnicity in physics, Brewe *et al.* [4] found that, while all students learned significantly more in courses with student-centered pedagogies than in courses with lecture-based instruction, gender differences increased from pre to posttest on conceptual inventories in both types of courses. However, Brewe *et al.* [4] also found that differences between majority and ethnic minority students did not increase in courses that used student-centered pedagogies while they did increase in courses that used lecture-based instruction.

The Learning Assistant (LA) model is one method for supporting the adoption and dissemination of student-centered instructional strategies in college science courses. An LA is an undergraduate student who are guided by course instructors and a special pedagogy course to facilitate discussions among groups of students in a variety of classroom settings that encourage student engagement and responsibility for learning. Over the last 10 years, the LA model has spread from a handful of physics and astronomy courses in 3 institutions to a range of STEM courses across 70+ institutions [5]. The broad dissemination of the model has led the LA-using institutions to create a support network called the LA Alliance. The LA Alliance has made it possible to measure the impact of LAs across institutional contexts.

LAs are associated with improved student performance in physics courses at multiple institutions [6, 7]. The impacts of LAs on equity is less clear. Kost-Smith *et al.* [8, 9] found that gender differences in concept inventory scores increased from pre to posttest. Their study, however, focused on a single institution that has highly effective physics courses and may not represent the larger set of LA-using courses.

The emergence of the LA Alliance and subsequent large-scale data collection using the LASSO platform enabled researchers to collect data from large enough samples to make reliable claims about the impact of LAs on physics students from underserved backgrounds. Using the first semester of data collected using LASSO, Van Dusen et al. [10] performed an exploratory analysis of the impacts of LAs across the STEM disciplines (1,645 students in 15 courses). In their analyses, Van Dusen et al. found evidence of persistent gender and racial inequities in LA-supported courses. Van Dusen et al. [11] performed a follow-up study examining the impact of LA-supported environments on first and second semester physics students from dominant and underserved backgrounds (2,868 students in 67 courses). Students from dominant backgrounds were defined as white or Asian, non-Hispanic, and male. They found that in LA-supported courses underserved students had larger shifts in their knowledge, measured as an effect size using Cohens *d*, than dominant students. In courses without LAs, however, they found the opposite trend; dominant students had larger effect sizes. The investigation presented in this publication builds on the Van Dusen *et al.*

[11] findings by focusing on first semester physics courses, increasing the statistical power, and creating more nuanced hierarchical linear models.

II. PURPOSE AND RESEARCH QUESTIONS

This study examines the role of LAs in supporting equity in college physics courses. To do this, we investigate the following research questions: (1) What gaps exist across gender and racial/ethnic student demographics in introductory physics courses? (2) How do student performance gaps in LA-supported and non-LA-supported introductory physics courses compare?

III. METHODS

Data collection: We accessed our large-scale, multi-institution data from the Learning About STEM Student Outcomes (LASSO) platform. The LASSO platform hosts, administers, scores, and analyzes student pre and posttest assessments online. Instructors download a report on their students' performance and have access to all of their students responses. Data from the courses are added to the LASSO database where they are anonymized, aggregated with similar courses, and made available to researchers with approved IRB protocols. Prior to taking the assessment on line, students are asked to complete a brief demographics questionnaire. For this study, we examined data from courses that used the Force Concept Inventory (FCI) [12] or Force and Motion Conceptual Evaluation (FMCE) [13]. We did not differentiate between the FCI and FMCE in the models we present because our preliminary analysis showed that doing so did not meaningfully change the model.

Data processing: We removed assessment scores for students if they took less than 5 minutes on the assessment or completed less than 80% of the questions. We removed entire courses if they had less than 40% student participation on either the pre or posttest. After cleaning the data we used hierarchical multiple imputation (HMI) with the hmi and mice packages in R to address missing data. HMI is a principled method for maximizing statistical power by addressing missing data while taking into account the structure of the data. HMI also can help ameliorate selection effects from participation rates skewing toward higher performing students [14]. HMI addresses missing data by (1) imputing each missing data point m times to create m complete data sets, (2) independently analyzing each data set, and (3) combining the m results using standardized methods [15]. After filtering but prior to running HMI our data was missing 15% of the pretest scores and 30% of the posttest scores. The analysis used 10 imputed datasets.

After cleaning and imputation, our dataset included 4,365 students from 93 courses. We used students self-reported demographic data to classify them using

TABLE I. Student demographics and performance

Demographics		N	Pre(S.D.)	Post(S.D.)	Gain
			Mean Score (%)		
Gender	Dominant	2,703	41.6(20.7)	61.9(23.7)	20.4
	Underserved	1,662	29.7(17.0)	49.7(23.4)	20.0
	Difference	-	-11.9	-12.3	-0.4
Race/ Ethnicity	Dominant	3,404	38.8(20.6)	59.9(24.1)	21.1
	Underserved	961	30.8(17.2)	47.9(22.9)	17.2
	Difference	-	-8.1	-12.0	-3.9

their gender (male = dominant gender; female or non-binary = underserved gender) and racial/ethnic identities (white or Asian and non-Hispanic = dominant race/ethnicity; neither white, Asian, or Hispanic = underserved race/ethnicity). In our sample, 1,662 (38.1%) of the students are of an underserved gender and 961 (22.0%) are of an underserved race/ethnicity. 56 (60.2%) of the courses were LA-supported and 79 (84.9%) of the courses used the FCI.

Data analysis: We calculated descriptive statistics to identify gaps between the average pre and posttest scores across student demographics. There were meaningful disparities in student pretest and posttest scores across student genders and races/ethnicities (Tab. 1). The gaps that students began the course with (11.9% for underserved gender and 8.1% for underserved race/ethnicity) are even wider by the posttest (12.3% for underserved gender and 12.0% for underserved race/ethnicity). These differences were within the range of gap sizes measured by Brewe *et al.* [4] for gender and for race.

To identify gaps in student performance, we used the HLM 7 software to create models that take the structure of data into account. Specifically, we developed 2-level Hierarchical Linear Models (HLM) that nest student data within course data. Our HLM models allowed us to quantify the interaction effect between a course being LA-supported and student demographic data while accounting for inherent and unknown course-level variations (e.g. the time of day of a class, student majors, and instructor backgrounds can lead to unforeseeable differences in student performance).

We developed our HLM models through a series of incremental additions of variables. In this paper we show the results from three models with postscore as the outcome variable. Model 1 is the unconditional model with no predictor variables. Model 2 includes the student (level-1) variables (gender, race/ethnicity, and student prescore). Model 3 builds on Model 2 by including the course (level-2) variables (LA-Supported and class mean prescore) and is shown below. The level-1 equation includes a coefficient for the intercept (β_{0j}), for the underserved gender (β_{1j}), underserved race/ethnicity (β_{2j}), student prescore (β_{3j}), and for a random effects variable (r_{0j}). Each coefficient in level 1 has an associated level

2 equation. In the level 2 equation, the intercept is γ_{i0}, there is an associated coefficient (γ_{ij}) for each variable in the equation and u_{ij} represents the random effect.

Level-1 Equation

$$(\text{Postscore})_{ij} = \beta_{0j} + \beta_{1j} * (\text{Und.serv. Gender})_{ij} +$$
$$\beta_{2j} * (\text{Und.serv. Race/Ethnicity})_{ij} +$$
$$\beta_{3j} * (\text{Student Prescore})_{ij} + r_{ij}$$

Level 2 Equations

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{LA-Supported})_j) +$$
$$\gamma_{02} * (\text{Class Mean Prescore})_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} * (\text{LA-Supported})_j + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21} * (\text{LA-Supported})_j + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$

LA-support is not included in the level-2 equation for student prescore because the interaction between the variables are not of interest in our analysis. For ease of interpretation, student prescore is group mean centered, class mean prescore is grand mean centered, and all other variables are uncentered. We included prescores in the model because they are strong predictors of student performance and improved the model's fit. Since prescores are not the focus of this investigation we will not discuss them in our interpretation of the models.

We compared Model 1's level-1 (r) and level-2 intercept variance (u_{0j}) (Tab. 2) to calculate the Intra-class Correlation Coefficient (ICC). In our case, the ICC identifies what percentage of the differences in student performance is attributed to student features (gender, race/ethnicity, and student prescore) versus course features (LA-support and class mean prescore). The ICC shows that course-level features explain 29% of the variance in student performance and student-level features explain the remaining 71%. These percentages show that course features have a substantial effect on student performance and HLM is an appropriate method of analysis. The reduction in level-1 variance (r) from Model 1 to Model 2 (Tab. 2) shows that our student-level variables explain 27% of the within-class variance in student performance. The reduction in level-2 intercept variance (u_{0j}) from Model 1 to Model 3 shows that our final model explains 57% of the variance in mean performance across classes. The reduction in the variances associated with gender (u_{1j}) and race/ethnicity (u_{2j}) from Model 2 to Model 3 show that our final model explains 67% of the gender gap and 27% of the race/ethnicity gap across classes. The explained variance shows that Model 3 has strong explanatory power. As Model 3 is our most robust model, we will focus on it in our findings section.

IV. FINDINGS

Model 3 reliably (p≤0.001) identifies gaps in posttest scores across student demographics while controlling for

TABLE II. Hierarchical Linear Models

	Fixed Effects with Robust SE					
	Model 1		Model 2		Model 3	
	γ	p	γ	p	γ	p
Intercept β_{0j}						
Intercept γ_{00}	55.05	<0.001	57.70	<0.001	54.52	<0.001
LA-Sup. γ_{01}	-	-	-	-	5.50	0.003
Class Pre γ_{02}	-	-	-	-	0.94	<0.001
For Underserved Gender slope β_{1j}						
Intercept γ_{10}	-	-	-3.99	<0.001	-3.54	0.001
LA-Sup. γ_{11}	-	-	-	-	-0.25	0.862
For Underserved Race/Ethnicity slope β_2						
Intercept γ_{20}	-	-	-4.30	<0.001	-4.08	<0.001
LA-Sup. γ_{21}	-	-	-	-	-0.21	0.884
For Student Prescore slope β_3						
Intercept γ_{30}	-	-	0.52	<0.001	0.51	<0.001

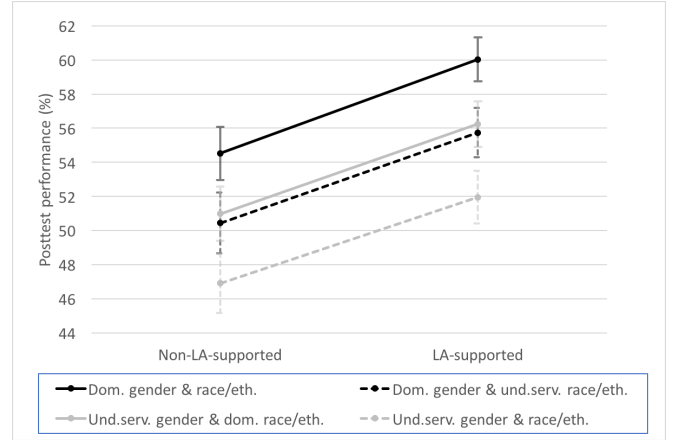|	Random Effect Variance		
	Model 1	Model 2	Model 3
Intercept u_{0j}	163.99	154.41	69.85
Gender u_{1j}	-	10.01	3.26
Race/Eth. u_{2j}	-	1.85	1.36
Prescore u_{3j}	-	0.01	0.01
Level-1 r_{ij}	405.62	297.42	297.62

FIG. 1. Predicted posttest scores (+/- 1 S.E.) controlling for pretest scores

student and class average pretest scores. In non-LA-supported courses, the model predicts that students from dominant and underserved genders who begin the class with the *same* pretest scores will have a difference in posttest scores of 3.5%. A similar gap (4.1%) emerges between students from dominant and underserved races/ethnicities. The model predicts that students in non-LA-supported courses who are underserved by gender *and* race/ethnicity will score 7.6% lower than their peers in dominant gender and race/ethnicity groups with equivalent pretest scores.

Model 3 shows student posttest scores were higher in LA-supported courses than traditional courses across all demographics. While students performed better overall

in LA-supported, the model shows the predicted performance gaps were not reliably (p>0.8) or meaningfully ($d\sim0.01$) smaller than in non-LA-supported courses. Figure 1 shows the predicted posttest score for students with average pretest scores across demographic groups in non-LA and LA-supported courses. The differences in scores between groups of students are very similar in both settings. While LAs were not associated with a reduction in the raw differences in average group posttest scores, they decreased the percentage in difference in student gains across groups. Courses with LAs have higher relative gains for students who are from underserved groups, 87% versus 80%, compared to the gain for students from dominant backgrounds.

average gain. Students who are underserved by gender *and* race/ethnicity miss out on over one third of the gain of their peers from dominant groups. The population of students who are underserved by both gender and race/ethnicity is small in physics and thus it is very difficult to investigate their performance with quantitative methods.

Contrary to the findings in our exploratory investigation that did not utilize nested models[16], raw inequities in student posttest scores were effectively constant across non-LA and LA-supported contexts. Because LAs were associated with improved outcomes for all students, LAs reduced the relative gaps in gains between student groups.

V. DISCUSSION

Differences in performance across genders in physics has been the focus of many investigations[3] and is well established. Research into the differences in performance across racial and ethnic lines in physics has received limited attention. Our results show that the inequities in performance by race and ethnicity are similar in size to the inequities by gender, indicating that these inequities deserve similar levels of attention. It is possible that the inequities by race and ethnicity have received little attention because underserved race/ethnicity students have so little representation in introductory physics courses that it has been very difficult for researchers to get reliable measures of these differences.

At first appearance the posttest difference for gender (3.5%) and race/ethnicity (4.1%) may seem like only a small difference. Given that the average improvement from pre to posttest for students with dominant in gender and race/ethnicity identities is approximately 20%, falling behind by 3.5% or 4.1% over the course of a semester represents missing out on nearly a fifth of the

VI. LIMITATIONS AND FUTURE WORK

This investigation identified consistent, reliable, and meaningful inequities in student performance in introductory physics courses. While our findings showed no significant differences in the gaps within classroom contexts, it is unclear how representative our non-LA-supported course data is of introductory physics courses more broadly. The LASSO platform has been primarily promoted to faculty in the LA Alliance, which likely skewed the courses to be ones that use research-based pedagogical practices whether they were LA-supported or not. Thus, the inequities that are in the courses without LAs may not be representative of inequities in traditional lecture-based courses. We expect that increased adoption of the LASSO platform will improve the generalizability of our findings. In our future work we will use a meta-analysis of published results to further inform our analysis.

[1] National Research Council, Tech. Rep. (2011).
[2] NRC, *DBER report*, Tech. Rep. (2012).
[3] E. Brewe and V. Sawtelle, Phys. Rev. PER (2016).
[4] E. Brewe, V. Sawtelle, L. H. Kramer, G. O'Brien, I. Rodriguez, and P. Pamelá, Phys. Rev. ST - PER (2010).
[5] V. Otero, L. Langdon, I. Her Many Horses, M. Oatley, and B. Van Dusen, in *Env. the Future of Undergrad. STEM Ed. Conf.* (2016).
[6] S. Pollock and N. Finkelstein, PERC Proc. 2012 (2012).
[7] R. M. Goertzen, E. Brewe, Wells L. Kramer, L H., and D. Jones, Phys. Rev. ST - PER (2011).
[8] L. Kost-Smith, S. Pollock, N. Finkelstein, G. Cohen, T. Ito, A. Miyake, C. Singh, M. Sabella, and S. Rebello, PERC Proc. (2010).
[9] L. Kost-Smith, S. Pollock, N. Finkelstein, G. Cohen, T. Ito, and A. Miyake, PERC Proc. (2012).
[10] B. Van Dusen, L. Langdon, and V. Otero, in *PERC Proc.* (2015).
[11] B. Van Dusen, J.S. White, and E. Roualdes, in *PERC Proc.* (2016).
[12] D. Hestenes, M. Wells, and G. Swackhamer, TPT (1992).
[13] R. Thornton and D. Sokoloff, Am. J. Phys. (1998).
[14] M. Jariwala, J. Nissen, X. Herrera, E. Close, and B. Van Dusen, in *PERC Proceedings* (in press).
[15] J. Drechsler, Jour. of Ed. and Beh. Stat. (2015).
[16] J. S. White, B. Van Dusen, and E. Roualdes, in *PERC Proc.* (2016).