

## Does Instructor Gender Matter for Student Performance in Introductory Physics?

Asim Ozmetin, Matthew Dew, and Tatiana Erukhimova  
*Texas A&M University, College Station, Texas 77843, USA*

Jonathan Perry  
*University of Texas at Austin, Austin, Texas 78712, USA*

A small number of prior studies have investigated whether instructor gender has a measurable impact on student course grades, though to our best knowledge no study has specifically examined physics courses. We explored the effect of instructor gender on student performance in calculus-based introductory physics courses at a large land-grant university. The data set included the scores on midterm, final exams, and final letter grades for more than 8200 students taught by 11 male and 4 female instructors between 2007-2018. We found no persistent correlation between student gendered performance and their instructor gender. While statistically significant differences in student performance were found for particular exams or in final letter grades, the effect sizes were weak or small. When comparing male and female instructors, we found that the instructor gender had a larger effect on female students than on their male classmates.

## I. INTRODUCTION

In recent years, physics education studies have found a consistent performance gap between male and female students on conceptual assessments [1–6], though this difference has been less consistent when it comes to differences in course grades [2, 3, 6–14]. Female students often find themselves to be a minority in their physics classes, and are likely to be subject to the pressures of societal biases and gender stereotype threats [15, 16]. It has been argued by prior studies that this stereotype threat can impact the performance of female students in their introductory courses [17, 18]. While some studies have reported on the reduction or elimination of gendered differences in undergraduate physics through specific interventions, such as values affirmation [19], or instructional strategies [20, 21], other groups have found no effect on these differences when applying selected pedagogies or when controlling for prior factors [1, 6, 22, 23].

A study from Lockwood et al. suggested that successful female role models, who were at a more advanced career stage, may serve as a proxy for students determining their own potential and may undermine gender stereotypes [24]. Stout et al. argued that female students who came into contact with experts of the same gender experienced enhanced STEM identity and self-efficacy, and improved motivation to pursue STEM careers [25]. A large-scale study of nearly 54000 students [26] considered how having a female faculty member in an introductory course affects the likelihood that a female student will take additional credit hours or major in a particular subject. The study suggested the existence of a positive impact of female instructors in some disciplines, but not in physics. A few studies have explored whether students' performance differences in STEM disciplines might be statistically impacted by the gender of their instructor. One such study conducted at the United States Air Force Academy found a significant effect on female students suggesting that the gender gap in course grades was eliminated when high-performing female students were assigned to female professors in mandatory introductory math and science courses [27]. Other studies aggregating data from multiple disciplines found small or no effects of instructor gender on students' course grades [28, 29]. To our best knowledge, there were no prior studies of the instructor gender effect that would focus on physics classes.

This work explores the intersection of instructor gender and student performance specifically for calculus-based introductory physics courses. It follows the footsteps of our recent publication that studied performance differences based on student gender in both calculus- and algebra-based introductory physics courses at a large land-grant university [14]. We expanded on this study with primary focus on instructor gender. We studied whether the existence and size of performance difference between male and female students depends on instructor gender while also exploring if there is a measurable difference in student performance depending on whether their instructor has the same gender.

## II. DATA ANALYSIS

We started with a data set previously gathered from faculty who taught one or both of a two-semester calculus-based introductory physics course sequence at a large, land-grant university between 2007-2018 [14]. The data set comprised more than 8200 engineering, math, and physical science majors from more than 75 lecture sections from both a first semester mechanics course and a second semester electricity and magnetism course during this period. The data included student name, instructor gender, instructor identifier, course year, exam grades, and final letter grades. In both courses there were three midterm exams and a final exam (labeled as exam 4 later on). Exams typically comprised 60%-90% of a student's total grade. Data were collected from 11 male instructors and 4 female instructors. Student first names were used to identify gender using GenderizeIO [30]. This tool returns a probability of gender for each first name based on census data and tends to be fairly robust with the exception of east Asian names [31]. Only student data with a 90% probability of being associated with a particular gender were kept which removed approximately 17% of the original data set.

All data analysis was done using Python programming language and its broadly accepted packages. We converted student final letter grades into their corresponding numeric GPA equivalents; F = 0, D = 1, C = 2, B = 3, A = 4. To analyze student performance across multiple years and instructors all raw numerical exam scores were mapped to a new distribution using a z score transformation,

$$z = \frac{x_i - \bar{x}}{\sigma}.$$

where  $x_i$  is the raw numerical score for each student,  $\bar{x}$  is the average score for that course's assessment, and  $\sigma$  is the standard deviation on that assessment. This new distribution was separated into student and instructor gender combinations. We separated the data set into four groups based on both instructor gender and student gender and analyzed these groups to examine difference in student performance on exams and course grades. In one case we inspected data separated by instructor gender, while in the other case we inspected data separated by student gender.

Groups were compared to one another using a two-tailed independent  $t$  test for statistically significant differences ( $p < 0.05$ ). Assumptions for the  $t$  test include homogeneity of variance as well as normality. We tested homogeneity of variance using Levene's test and found that nearly all of the groups being hypothesis tested had a uniform variance. Likewise, normality of the data was examined using the Shapiro-Wilk test and was found to fail. However, with sample sizes above one thousand the  $t$  test remains to be the most robust measure despite the normality assumption not being met [32]. Effect sizes were calculated using Cohen's  $d$  with a Hedges' correction [33]. Positive values correspond to higher averages for male students with negative values corresponding to higher averages for female students. The effect sizes were consid-

ered to be weak for  $|d| < 0.2$  and small for  $0.2 < |d| < 0.5$ .

### III. RESULTS & DISCUSSION

First we investigated the student gendered performance difference on exams, within same instructor gender. Table I shows the differences between male and female students average exam z scores for all male instructors (left column) and all female instructors (right column) as well as corresponding  $t$  and  $p$  values. For students of male instructors, which included  $N_{male} = 3942$  male students and  $N_{female} = 1202$  female students, statistically significant differences were found on the first, third, and fourth exams. The corresponding effect sizes,  $d = 0.228, 0.079, 0.144$ , were small, weak, and weak respectively. For students of female instructors,  $N_{male} = 2397$  and  $N_{female} = 701$ , statistically significant difference between male and female students was found on the first exam with a weak effect size,  $d = 0.013$ .

TABLE I. Mean exam z score differences,  $\Delta$ , between male and female students and corresponding  $t$  test results and p-values, separated by *instructor* gender. Statistically significant differences have been noted with an asterisk(\*).

	Male instructor			Female instructor		
	$\Delta(M - F)$	$t$	$p$	$\Delta(M - F)$	$t$	$p$
Exam 1	0.227	6.914	<0.0005*	0.121	2.809	0.005*
Exam 2	0.026	0.790	0.429	0.043	1.009	0.313
Exam 3	0.079	2.395	0.017*	0.010	0.242	0.809
Exam 4	0.144	4.382	<0.0005*	0.013	0.306	0.760

We also examined differences for final letter grades. For female instructors, Fig. 1, there was a statistically significant difference in final letter grades, with female students exhibiting a slightly higher average grade. This difference had a weak effect size ( $d = 0.084$ ). For male instructors, Fig. 2, there was no statistically significant difference in final letter grades based on student gender. From this we can observe that female students exhibit a small but measurably higher performance compared to male students with female instructors, while there is no statistically measureable difference for male instructors.

Next we consider a similar comparison from an alternative perspective, separating data based on student gender: we examined whether students of a certain gender performed statistically different with male and female instructors. Table II shows averaged exam z scores differences between male and female instructors for all male students (left column) and all female students (right column) with corresponding  $t$  and  $p$  values. There were  $N_{male} = 3942$  male students of male instructors and  $N_{female} = 2397$  male students of female instructors. We observed no statistically significant differences for male students when it comes to instructor gender. As for female students,  $N_{male} = 1202$  and  $N_{female} = 701$ , final exam grades showed a statistically significant difference between male and female instructors, favoring the female in-

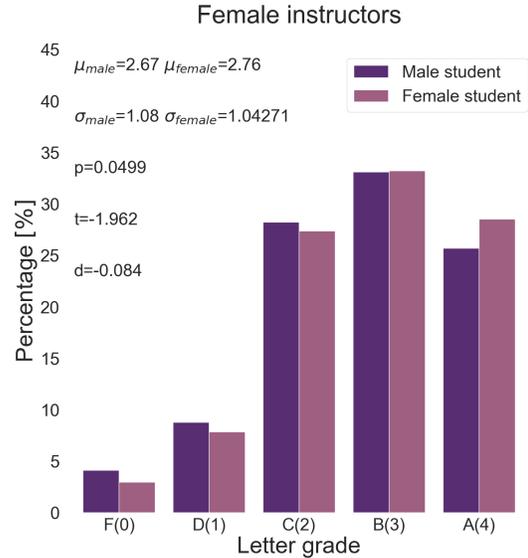


FIG. 1. Grade distributions for students with a female instructor, separated by student gender. Population averages ( $\mu$ ), standard deviations ( $\sigma$ ),  $t$  test results ( $t$ ), and effect size of the difference ( $d$ ) are included.

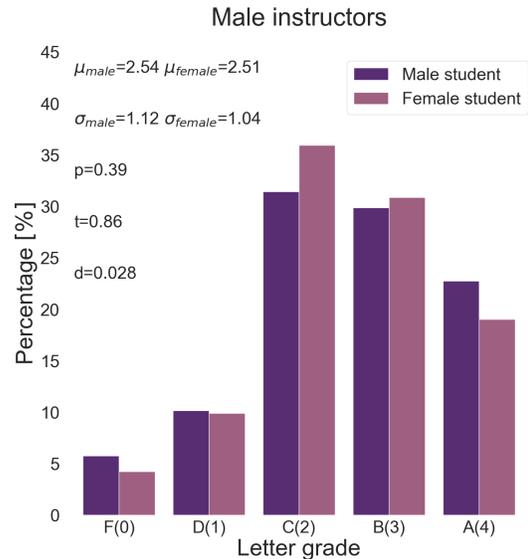


FIG. 2. Grade distributions for students with a male instructor, separated by student gender. Population averages ( $\mu$ ), standard deviations ( $\sigma$ ),  $t$  test results ( $t$ ), and effect size of the difference ( $d$ ) are included.

structors. Nonetheless, the effects size for this hypothesis testing,  $d = -0.106$ , was likewise weak.

As with the previous groups, we examined differences for final letter grades for groups based on student gender. For

TABLE II. Mean exam z score differences,  $\Delta$ , between male and female instructors and corresponding  $t$  test results and p-values, separated by *student* gender. Statistically significant differences have been noted with an asterisk(\*).

	Male student			Female student		
	$\Delta(M - F)$	$t$	$p$	$\Delta(M - F)$	$t$	$p$
Exam 1	0.026	0.999	0.318	-0.081	-1.684	0.092
Exam 2	-0.004	-0.143	0.887	0.014	0.295	0.768
Exam 3	0.016	0.614	0.539	-0.052	-1.144	0.253
Exam 4	0.031	1.171	0.242	-0.100	-2.219	0.027*

female students, Fig. 3 there was a statistically significant difference in final letter grades, with higher grades being associated with female instructors. This difference had a small effect size of  $d = -0.248$ . For male students, Fig 4 there was also a statistically significant difference in final letter grades, with higher grades again being associated with female instructors. This difference had a weak effect size of  $d = -0.125$ . It is interesting to note that both male and female students earned higher grades in courses taught by female instructors.

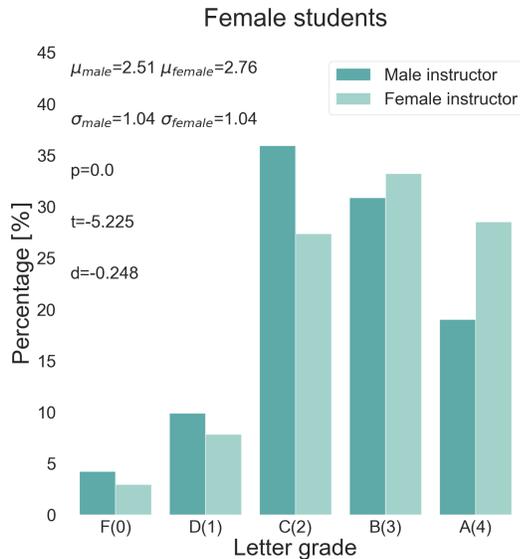


FIG. 3. Grade distributions for female students separated by instructor gender. Population averages ( $\mu$ ), standard deviations ( $\sigma$ ),  $t$  test results ( $t$ ), and effect size of the difference ( $d$ ) are included.

Prior studies that examined the impact of instructor gender on student performance incorporated data across multiple STEM disciplines, though none specifically looked at physics. However, we found that our results of instructor gender having only a minimal relation to final course grades were in general agreement with several of these multi-disciplinary studies [28, 29]. For example, for male instructors, Fig. 2, there were no statistically significant differences between final grades for male and female students. We observed a sta-

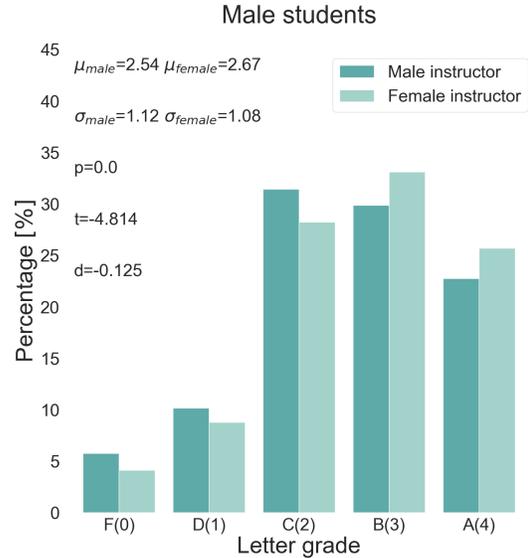


FIG. 4. Grade distributions for male students separated by instructor gender. Population averages ( $\mu$ ), standard deviations ( $\sigma$ ),  $t$  test results ( $t$ ), and effect size of the difference ( $d$ ) are included.

tistically significant difference for female instructors, Fig. 1, with a weak effect size. The results presented on gendered performance on exams, Table I, mirrors results from our prior study [14] with fewer statistically significant differences for female instructors compared to male instructors.

Our results from grouping the data by student gender, echo results from the Air Force Academy study [27], with instructor gender being a more significant factor for female students than for male students. The exam performance results, Table II, display only one instance of a statistically significant difference which occurs for female students on the fourth exams. For male students, no statistically significant differences were seen to occur on any exams between male and female instructors, even though the differences existed in the final letter grades (with a weak effect size). This is in agreement with Lockwood et al. who observed that the gender of a role model, who could be an instructor, had no impact on male students [24].

#### IV. CONCLUSIONS

We have analyzed a data set of more than 8200 students' exam scores and final course grades over more than a decade from instructors who taught calculus-based introductory physics courses at a large, land-grant university. Data from a first semester mechanics course and a second semester electricity and magnetism course were collected from 4 female instructors and 11 male instructors. We examined whether instructor gender had a measurable impact on stu-

dents' exams grades throughout a semester and final course grades. First, we analyzed data pooled by student gender exploring performance difference between male and female students with male and female instructors. We also inspected data grouped by instructor gender to determine if there were differences in student performance with an instructor of the same gender. Our analysis did not reveal any significant, persistent correlation between student gendered performance and their instructor gender. In all cases where we found statistically significant differences in student performance on particular exams or in final letter grades, the effect sizes were weak or small. Similar to our recent study [14], we found fewer statistically significant differences in exam performance between male and female students for female instructors compared to male instructors. A notable result is that when comparing male and female instructors, we found that instructor gender had a larger effect on female students than on their male classmates.

To our best knowledge, this work and our recently published paper [14] are first steps in investigating whether instructor gender has a measurable impact on student performance in physics courses. Though the results presented here have the advantage of a relatively large data set of exam scores and final course grades collected over a decade, we

must note several limitations. The data collected was provided by a modest number of instructors, with female instructors representing just four out of fifteen. Female instructors taught 37.5% of the total population, however, it should be noted that one female instructor taught approximately half of those students. Only course level data were collected for this study. Also, student gender was determined by probability as opposed to using university level data.

Future studies would benefit from the inclusion of a larger number of female instructors, as well as considering prior student preparation. It would be also important to analyze data from algebra-based introductory physics classes with larger number of female students enrolled. More work is needed to make a more definitive conclusion on the role of gender on student performance in the introductory course sequences.

## V. ACKNOWLEDGEMENTS

This work was partially supported by the Texas A&M University College of Science. We are grateful to the faculty from the department of physics and astronomy at Texas A&M University who provided data for this study.

- 
- [1] A. Madsen, S. B. McKagan, and E. C. Sayre, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020121 (2013), URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.9.020121>.
  - [2] J. Docktor and K. Heller, *AIP Conference Proceedings* **1064**, 15 (2008), URL <https://aip.scitation.org/doi/abs/10.1063/1.3021243>.
  - [3] S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch, and N. R. Walet, *European Journal of Physics* **34**, 421 (2013), ISSN 1361-6404, URL <http://dx.doi.org/10.1088/0143-0807/34/2/421>.
  - [4] A. Traxler, R. Henderson, J. Stewart, G. Stewart, A. Papak, and R. Lindell, *Phys. Rev. Phys. Educ. Res.* **14**, 010103 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.14.010103>.
  - [5] R. Henderson, P. Miller, J. Stewart, A. Traxler, and R. Lindell, *Phys. Rev. Phys. Educ. Res.* **14**, 020103 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.14.020103>.
  - [6] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010107 (2007), URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.3.010107>.
  - [7] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009), URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.5.010101>.
  - [8] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, *Science* **330**, 1234 (2010), ISSN 0036-8075, URL <https://science.sciencemag.org/content/330/6008/1234>.
  - [9] R. L. Matz, B. P. Koester, S. Fiorini, G. Grom, L. Shepard, C. G. Stangor, B. Weiner, and T. A. McKay, *AERA Open* **3**, 2332858417743754 (2017), URL <https://doi.org/10.1177/2332858417743754>.
  - [10] E. Brewster, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010106 (2010), URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.6.010106>.
  - [11] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, and A. Traxler, *Phys. Rev. Phys. Educ. Res.* **13**, 020114 (2017), URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.13.020114>.
  - [12] S. Lauer, J. Momsen, E. Offerdahl, M. Kryjevskaja, W. Christensen, and L. Montplaisir, *CBE Life Sciences Education* **12**, 30 (2013), pMID: 23463226, URL <https://doi.org/10.1187/cbe.12-08-0133>.
  - [13] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020112 (2010), URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.6.020112>.
  - [14] M. Dew, J. Perry, L. Ford, W. Bassichis, and T. Erukhimova, *Phys. Rev. Phys. Educ. Res.* **17**, 010106 (2021), URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.17.010106>.
  - [15] L. McCullough, *The Physics Teacher* **40**, 86 (2002), URL <https://doi.org/10.1119/1.1457312>.
  - [16] C. Nord, S. Roey, R. Perkins, M. Lyons, N. Lemanski, Y. Tamir, J. Brown, J. Schuknecht, and K. Herrold (2011), retrieved from <https://nces.ed.gov/nationsreportcard/pdf/studies/2011462.pdf>.
  - [17] I. Goodman, C. Cunningham, C. Lachapelle, M. Thompson, K. Bittinger, R. Brennan, and M. Delci (2002).
  - [18] L. Bian, S.-J. Leslie, and A. Cimpan, *Science* **355**, 389 (2017), ISSN 0036-8075, URL <https://science.sciencemag.org/content/355/6323/389>.
  - [19] M. Lorenzo, C. H. Crouch, and E. Mazur, *American Journal of Physics* **74**, 118 (2006), URL <https://doi.org/10.1119/>

- 1.2162549.
- [20] T. Espinosa, K. Miller, I. Araujo, and E. Mazur, Phys. Rev. Phys. Educ. Res. **15**, 010132 (2019), URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.15.010132>.
- [21] I. Rodriguez, G. Potvin, and L. H. Kramer, Phys. Rev. Phys. Educ. Res. **12**, 020118 (2016), URL <https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.020118>.
- [22] N. I. Karim, A. Maries, and C. Singh, European Journal of Physics **39**, 025701 (2018).
- [23] M. J. Cahill, K. M. Hynes, R. Trousil, L. A. Brooks, M. A. McDaniel, M. Repice, J. Zhao, and R. F. Frey, Phys. Rev. ST Phys. Educ. Res. **10**, 020101 (2014), URL <https://link.aps.org/doi/10.1103/PhysRevSTPER.10.020101>.
- [24] P. Lockwood, Psychology of Women Quarterly **30**, 36 (2006), URL <https://doi.org/10.1111/j.1471-6402.2006.00260.x>.
- [25] J. Stout, N. Dasgupta, and M. Hunsinger, Soc. Psychol **100**, 255 (2011).
- [26] E. P. Bettinger and B. T. Long, American Economic Review **95**, 152 (2005).
- [27] S. E. Carrell, M. E. Page, and J. E. West, The Quarterly Journal of Economics **125**, 1101 (2010), ISSN 0033-5533, URL <https://doi.org/10.1162/qjec.2010.125.3.1101>.
- [28] F. Hoffmann and P. Oreopoulos, Journal of Human Resources **44**, 479 (2009).
- [29] S. M. Solanki and D. Xu, American Educational Research Journal **55**, 801 (2018).
- [30] C. Strömngren, *Genderizeio* (2013), available at <https://genderize.io/>, URL <https://genderize.io/>.
- [31] J. Huang, A. J. Gates, R. Sinatra, and A. L. Barabási, Proceedings of the National Academy of Sciences **117**, 4609 (2020).
- [32] M. W. Fagerland, BMC Med Res Methodol **12**, 78 (2012).
- [33] L. Cohen, L. Manion, and K. Morrison, *Research Methods in Education* (Routledge, New York City, New York, 2018).