Gendered Performance Differences in Introductory Physics: A Study from a Large Land-Grant University

Matthew Dew,1 Jonathan Perry,2 Lewis Ford,1 William Bassichis,1 and Tatiana Erukhimova1

1Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843
2Department of Physics, University of Texas, Austin, TX 78712

(Dated: September 22, 2020)

Studies examining gender differences in introductory physics show a consensus when it comes to a gender gap on conceptual assessments; however, the story is not as clear when it comes to differences in gendered performance on exams. This study examined gendered differences in student performance in two introductory physics course sequences to determine whether they were persistent throughout each course. The population for this study included more than 10,000 students enrolled in algebra- and calculus-based introductory physics courses between spring 2007 and spring 2019. We found a small but statistically significant difference in final letter grades for only one out of four courses, algebra-based mechanics. By looking at midterm exam grades, statistically significant differences were noted for some exams in three out of four courses, with algebra-based electricity and magnetism being the exception. In all statistically significant cases, the effect size was small or weak, indicating that performance on exams and final letter grades was not strongly dependent on gender. Additionally, a questionnaire was administered in fall 2019 to more than 1,600 students in both introductory sequences to measure students’ perceptions of performance, class contributions, and inclusion. We observed differences between students’ perception of their performance and contribution when grouped by gender, but no difference on perception of inclusion.

I. INTRODUCTION

Over the last half century, the number of US students majoring in STEM fields has more than doubled [1]. As this enrollment has increased, so has the attention paid to who is obtaining degrees across different disciplines, particularly when it comes to underrepresented groups. While some STEM disciplines, such as biology, have relative parity between males and females attaining degrees, other disciplines have a persisting gender gap [2]. The National Center for Science and Engineering Statistics found that in 2016, women earned 20.9% of all engineering bachelor’s degrees and 19.3% of all physics degrees [3].

Out of all STEM disciplines, physics is often considered to be the field that is the least welcoming for women to join [4, 5]. Even for students not majoring in physics, STEM majors have to take physics as part of their academic program. For physics and engineering majors, introductory physics courses are among their early experiences in college. Such experiences can be crucial for student success within their majors [6].

The importance of such experiences is especially true for female students, as many of them leave physical science and engineering tracks during the first two years of college [7]. Female students are likely to be under the pressure of gender stereotypes and societal biases [8, 9], and they often find themselves underrepresented in their physics classes. Some authors argue that stereotype threat influences female student performance in introductory physics classes [10, 11] and that use of an intervention based on value affirmation can help improve the situation [12]. Perhaps related to stereotyping, the atmosphere in physics classrooms can influence female students’ physics self-efficacy, self-identity and motivation; all of which can have an impact on student success and retention [13–17]. A number of studies have reported on the difference in physics self-efficacy between male and female students, including courses which use research-supported instructional methods [18–22]. As an example, Marshman et al. reported that female students had significantly lower self-efficacy than male students throughout a two-semester introductory physics course sequence [23]. They go on to note that the physics self-efficacy of female students was negatively impacted by both traditional instruction courses and flipped classroom courses.

A vast literature exists that explores the gendered differences in student performance on concept inventory tests in introductory physics courses. The majority of existing studies report a persistent gender gap with males performing significantly better than females on introductory mechanics concept inventory assessments [24–29], with some authors arguing that removing gender-biased context can reduce the gap [27, 28, 30]. The gender gap has been found in conceptual inventories of electromagnetism as well, although to a lesser extent and with more variation across studies [24–28, 31–32].

The results of prior studies on the gendered differences in student performance based on course grades and examinations are less consistent: while a number of studies indicate that male students outperform female students on the exams and course grades [12, 33–34], other groups found no significant gendered difference in student performance [25, 26, 29, 31, 32, 35, 36]. One study, comprised of 4,000 students across 7 semesters at the University of Colorado Boulder, reported a small but significant difference in course grades, correlated with differences in back-
ground factors for males and females [33]. Factors beyond the course, including prior knowledge, math background, and attitudes towards science, have been seen to correlate with gendered differences in performance [33, 37]. One study of an electricity and magnetism course by Andersson and Johansson argues that the gendered difference in course grades disappears when controlled for the program in which a student is enrolled [35]. Tai and Sadler show that females outperformed males in algebra-based courses while males outperformed females with the same background in calculus-based courses [39]. Several studies performed on a large number of students taking the introductory physics classes report no significant gendered difference in student performance on course exams and course grades but a gender gap in concept inventories [25, 26, 31]. Some studies of gendered differences in undergraduate physics have reported reduction or elimination through the use of carefully selected instructional strategies in introductory physics [21, 30, 41]. However, other groups have found no effect of applying selected pedagogies or controlling the prior knowledge factors on gendered performance [24, 29, 42, 43].

This work focused on expanding the studies of academic gendered differences through large enrollment courses at Texas A&M University (TAMU). This is a large, land-grant institution, which yearly serves more than 20,000 undergraduate STEM majors across multiple colleges [44]. STEM majors at TAMU complete their introductory physics sequence through either calculus-based courses (e.g., physics, engineering, chemistry, math) or algebra-based courses (e.g., life science majors, pre-meds, and environmental science). The engineering program comprises more than half of all STEM majors at TAMU, so the calculus-based sequence enrollment is much larger than the algebra-based sequence. Students demographics, academic goals, and attitudes towards physics may differ significantly between calculus-based and algebra-based courses. Most students in the calculus-based courses were in their freshman year, i.e., right after high school. The algebra-based courses are typically taken by upper-level students in their sophomore or senior year, who do not have physics or physics-related disciplines as the focus of their studies and careers. Furthermore, there is a much larger proportion of women in algebra-based courses.

Our study aimed to examine the evolution of gendered differences in algebra- and calculus-based introductory physics courses at TAMU by looking at both exams and final letter grades. We collected and analyzed data on students test performances since 2007. While working with our data set, we also wanted to take a snapshot of current students’ feelings to see how their perceptions aligned with historic performance. A short questionnaire was distributed to all students enrolled in calculus-based and algebra-based mechanics and electricity and magnetism in the fall of 2019.

### II. METHODS

From here forward, “significant” will be used as shorthand for “statistically significant”. Statistical significance was taken to be at $p < 0.05$. In addition, tables will use “Mech.” and “E&M” for mechanics and electricity and magnetism, respectively; “Alg.” or “Calc.” stand for algebra-based or calculus-based.

#### A. Course Data

To examine the gendered student performance within introductory courses, course level data was requested from faculty who taught one or more of these courses since 2007. Participating instructors provided students’ first names, numerical scores for all midterm and final exams, and the final letter grade for the course. After collecting data from faculty, a database of approximately 13,000 students was obtained. This database contains information for students enrolled in the algebra-based sequence between 2011-2019 and the calculus-based sequence between 2007-2017. This study was structured in such a way that the only data collected were the course-level information provided by faculty. For this reason, connecting outcomes to non-academic factors was not possible with this study.

Since course-level data only included student names, gender was identified using an online tool, GenderizeIO [45]. This application program interface returns a probability of gender based on the input of a first name. Gender probability was considered identifiable for this study if it was 90% or higher. This percentage was chosen as it allowed reasonable certainty of gender without drastically reducing the size of our data set. This cut eliminated about 17% of our raw sample. The number of students identified as male or female from each of the four introductory courses examined in this study is shown in Table I.

| Course          | Total Number | Male  | Female |
|-----------------|--------------|-------|--------|
| Mech. Alg.      | 1,267        | 44.4% | 55.6%  |
| E&M Alg.        | 999          | 44.6% | 55.4%  |
| Mech. Calc.     | 5,449        | 74.9% | 25.1%  |
| E&M Calc.       | 2,793        | 80.8% | 19.2%  |

#### B. Comparing Grades

Differences based on gender were examined by looking at students’ final course grades and their scores on the
midterm and final semester exams. Differences in populations were examined using t-tests between transformed data, as well as analysis of variance (ANOVA) applied to raw scores. Comparisons were made based on student gender, instructor, and year in the course. Some instructors gave multiple years of data, so these criteria allowed for individual lecture section distributions to be examined against each other. Though the exam distributions can be skewed for both raw scores and z-scores (see Figure 1 for raw Exam 1 scores from calculus-based mechanics), t-tests were the most appropriate statistical analysis due to the large sample size from each course [46]. Effect sizes were calculated using Cohen’s d with a Hedges correction [47]. We consider $d<0.2$ to be weak, $0.2<d<0.5$ to be small, $0.5<d<0.8$ to be medium, and $d>0.8$ to be large effect sizes.

To look specifically at the relation between gender and exam performance, raw scores from individual lecture sections were mapped to new distributions using a z-score transformation. A z-score takes a raw numerical score ($x_i$), subtracts the average ($\bar{x}$), and scales by the standard deviation ($\sigma$), according to the relation:

$$z = \frac{x_i - \bar{x}}{\sigma}.$$  (1)

A positive z-score indicates how much higher a raw score was compared to the average in units of standard deviation. A negative z-score indicates the same but for a raw score below the average [47]. This transformation of scores was performed for a more even comparison of exam distributions across multiple years and instructors. For instance, Professor A teaching a course in year X might have a higher average and smaller deviation than when the same instructor teaches the same course in year Y. As an illustration of the z-score transformation, we can look at raw exam scores from calculus-based mechanics. In 2007, the mean score was 58 points and the standard deviation was 21 points. Students scoring a 58 were mapped to a z-score of 0, while students scoring a 37 were mapped to a z-score of -1. This was done for all students, using the individual lecture section averages and standard deviations.

Raw scores were used to examine individual lecture sections. When comparing across lecture sections, raw scores were transformed into z-scores so distributions may be more adequately and fairly compared to one another. Final course grades were treated on a 4 point scale (A-F) with no plus or minus grades per TAMU’s grading policy.

C. Student Perceptions

In fall 2019, a short anonymous questionnaire was administered to explore how students felt about their performance and inclusion. This questionnaire was given to compare how students’ perceptions of their performance aligned with historic data.

III. ANALYSIS AND RESULTS

Differential performance for male and female students was examined for all exams and final course grades for
four introductory physics courses. Results are separated by courses in the algebra-based sequence and the calculus-based sequence. The calculus-based sequence consists of three midterm exams administered throughout the semester with a comprehensive final (identified as Exam 4). The algebra-based sequence consists of four midterm exams administered throughout the semester with a comprehensive final (identified as Exam 5).

Questionnaire responses were converted into an ordinal 5-point scale, with higher numbers equating to more positive responses. That is, better feelings of inclusion, greater performance, and stronger feelings of making valued contributions.

### TABLE II. Average final letter grades (and standard error) by gender for the algebra- and calculus-based introductory sequences, as well as the t-test and significance between these distributions.

|       | Male        | Female       | t-test | p   |
|-------|-------------|--------------|--------|-----|
| Mech. Alg. | 2.687 (0.043) | 2.839 (0.039) | -2.634 | 0.009 |
| E&M Alg.  | 2.935 (0.046) | 2.875 (0.041) | 0.967  | 0.334 |
| Mech. Calc. | 2.532 (0.018) | 2.517 (0.030) | 0.430  | 0.667 |
| E&M Calc.  | 2.596 (0.024) | 2.692 (0.047) | -1.786 | 0.074 |

### A. Calculus-Based Mechanics and E&M

For calculus-based mechanics, data were provided by 14 instructors, for a total of 49 lecture sections. Eleven male instructors provided data from 34 lecture sections, and three female instructors provided data from 15 lecture sections. As seen in Table II, there was no significant difference observed when a t-test was applied to final letter grades based on student gender for the pooled data from all instructors and sections.

Gendered performance on course exams were compared using t-tests on transformed data from all instructors, Table III. As a combined sample, male students score at least slightly higher compared to female students on all exams. In this case, none of these differences were significant.

As a validation of results found using t-tests on transformed data, three-way ANOVA was applied to raw scores for all exams from both calculus-based courses. Results were in agreement with t-tests applied to transformed data. Where significant differences were observed using t-tests, ANOVA showed gender to be a significant factor on its own or in combination with one or both of the other factors of professor and year. When examining individual lecture sections, significant differences due to gender were observed for less than 20% of lecture sections. Combined with the results above, we note a persistent gender difference in calculus-based mechanics on exams only for pooled data, producing no significant difference in final course grades. No gendered differences were noted in calculus-based E&M for either exams or final course grades.

### TABLE III. Average exam z-scores (and standard error) for calculus-based mechanics, as well as the t-test and significance between these distributions.

|       | Male        | Female       | t-test | p   |
|-------|-------------|--------------|--------|-----|
| Exam 1 | 0.060 (0.015) | -0.180 (0.027) | 7.780  | <0.001 |
| Exam 2 | 0.007 (0.016) | -0.020 (0.026) | 0.858  | 0.391 |
| Exam 3 | 0.016 (0.016) | -0.049 (0.026) | 2.099  | 0.036 |
| Exam 4 | 0.026 (0.016) | -0.077 (0.025) | 3.305  | 0.001 |

### B. Algebra-Based Mechanics and E&M

For algebra-based mechanics, data were provided by 4 instructors, covering a total of 13 lecture sections. Three male instructors provided data from 11 lecture sections, and one female instructor provided data from 2 lecture sections. Comparing the final letter grades based on student gender for pooled data from all instructors and sections showed a significant difference, Table IV. This significant difference is not observed across individual instructors, nor is it consistently observed for lecture sections. Only one lecture section exhibited a significant difference for letter grades.

### TABLE IV. Average exam z-scores (and standard error) for calculus-based E&M, as well as the t-test and significance between these distributions.

|       | Male        | Female       | t-test | p   |
|-------|-------------|--------------|--------|-----|
| Exam 1 | 0.013 (0.021) | -0.053 (0.042) | 1.360  | 0.174 |
| Exam 2 | 0.008 (0.021) | -0.034 (0.043) | 0.862  | 0.389 |
| Exam 3 | 0.005 (0.021) | -0.020 (0.042) | 0.502  | 0.616 |
| Exam 4 | 0.014 (0.021) | -0.057 (0.043) | 1.475  | 0.140 |
Gendered performance on course exams for transformed data from all instructors for algebra-based mechanics is shown in Table V. On the first exam of the semester, male students were observed to have a higher average score compared to female students, though the reverse is seen to be true for the other four exams of the course. Significant differences were observed for the third and fourth exams. The effect size for both of these exams is weak ($d < 0.2$).

For algebra-based E&M, data were provided by 2 instructors, covering a total of 8 lecture sections. Two male instructors provided all of the data for this course. No significant difference in final letter grades based on student gender for pooled data from all instructors and sections was observed for this course, Table II.

As with the algebra-based mechanics course, exams were compared using t-tests on transformed data from all instructors, Table VI. Male students were observed to have a higher average score on four out of five exams with female students outscoring male students on the third exam. A significant difference was observed only for the first exam. The effect size for this difference is weak ($d < 0.2$).

As a validation of results found using t-tests on transformed data, three-way ANOVA was applied to raw scores for all exams from both algebra-based courses. Results were in agreement with t-tests applied to transformed data. Where significant differences were observed using t-tests, ANOVA showed gender to be a significant factor on its own or in combination with one or both of the other factors of professor and year. When examining individual lecture section level data, significant differences due to gender were observed for less than 15% of lecture sections. Combined with the results above, we note a gender difference in algebra-based mechanics on the third and fourth exams only for pooled data, producing a small but significant difference in final course grades. A gender difference is observed in algebra-based E&M only for the first exam, producing no difference in final course grades.

TABLE VI. Average exam z-scores (and standard error) for algebra-based E&M, as well as the t-test and significance between these distributions.

|       | Male        | Female      | t-test | p  |
|-------|-------------|-------------|--------|----|
| Exam 1| 0.077 (0.047)| -0.062 (0.043)| 2.182  | 0.029|
| Exam 2| 0.053 (0.048)| -0.042 (0.042)| 1.492  | 0.136|
| Exam 3| -0.008 (0.045)| 0.007 (0.044)| -0.235 | 0.814|
| Exam 4| 0.023 (0.049)| -0.019 (0.041)| 0.653  | 0.514|
| Exam 5| 0.022 (0.047)| -0.018 (0.043)| 0.624  | 0.533|

C. The Effect of Instructor Gender

The impact of instructor gender on differences in student performance by gender was also examined. This analysis could not be applied to the algebra-based E&M course as all data was from male instructors. Data was separated into two groups by instructor gender, and comparisons were made based on student gender. Differences in average z-scores for each course exam are shown in Table VII, while comparisons between letter grades are shown in Table VIII.

For algebra-based mechanics, data were obtained from three male instructors for 11 lecture sections (N=1075) and from one female instructor for 2 lecture sections (N=192). Significant differences in student performance based on gender were observed during the course only for the fourth exam with male instructors. This difference has a weak effect size ($d < 0.2$). A significant difference in final letter grades was also observed for male instructors, with a weak effect size ($d < 0.2$).

For calculus-based mechanics, data were obtained from eleven male instructors for 34 lecture sections (N=4,227) and from three female instructors for 15 lecture sections (N=1,222). Significant differences in student performance based on gender were observed for instructors of both genders on the first midterm exam, with a small effect size for each ($0.2 < d < 0.4$). Other significant differences were noted for the third and fourth exams for male instructors only. These latter two differences have weak effect sizes ($d < 0.2$). No significant difference was noted in final letter grades when grouping students by instructor gender.

For calculus-based E&M, data were obtained from six male instructors for 17 lecture sections (N=1,876). Significant differences in student performance based on gender were observed for instructors of both genders on the first midterm exam, with a small effect size for each ($0.2 < d < 0.4$). A significant difference was also observed in letter grades for female instructors, with a weak effect size ($d < 0.2$).

We also analyzed the raw score data using ANOVA. Agreement on significance using t-tests and ANOVA was found for comparisons based on instructor gender except for two instances. These instances were calculus-based
mechanics on the first exam for female instructors and calculus-based E&M on the final exam for male instructors. Further examination of the impact of instructor gender was performed using Tukey HSD, which found significance in agreement with t-tests [47].

D. Students Perception Questionnaire

A short questionnaire consisting of three questions that were analyzed independently as well as demographic information was distributed to the students taking introductory physics classes in the fall of 2019, as described in Section II C. We received over 1,600 completed surveys with a response rate of 63%.

On the question about students perception of their performance,

2. “I believe that I performed ______ in this course.”

they were given five answer choices:

A. Well Below Average
B. Below Average
C. Average
D. Above Average
E. Well Above Average

We converted these answers into a 5-point scale with “Well Below Average” corresponding to a 1 and “Well Above Average” to a 5.

In the calculus-based mechanics course, male students rated their performance higher than their female classmates to a significant degree. Specifically, they rated themselves higher by a third of a point (See Table IX). We found that female students only rated their performance perception equally to male students in algebra-based mechanics (See Table IX). When compared to historic data, this is the one course examined where we found a statistically significant difference in final letter grades with female students outperforming male students.

For the other three courses, male students rated their performance as significantly higher. For both E&M courses, however, male students performed the same as female students at this point in the course according to historic data.

Next, we looked at how students rated their feelings of inclusion. More specifically, we asked the following,

1. “I felt included by my peers and instructors within this physics course.”

Finally, we asked students whether they felt that their contributions were valued.

3. “I felt that my contributions to discussions over physics material were valued during this course. This includes discussions both in class and outside of class but relating to completing assignments or preparing for exams.”

For both of these questions, the students were given the following answer choices:

A. Strongly Disagree
B. Disagree
C. Neither Agree nor Disagree
D. Agree
E. Strongly Agree

Similar to the question on performance perception, we converted these answers into a 5-point scale with “Strongly Disagree” corresponding to a 1 and “Strongly Agree” to a 5.

We found that despite a difference in performance perception, there was no statistically significant difference in feelings of inclusion for any of the courses (See Table X). That is, despite female students believing they were underperforming in three courses, they still believed they were being included equally.

When analyzing whether students felt that their contributions were valued, we found a statistically significant difference in algebra-based mechanics (See Table XI). In this course, female students rated valuations of their contributions about a fourth of a point higher on average. This corresponds to the only course where female students historically performed better on final letter grades. We found no significant differences in the other courses, indicating male and female students had similar feelings about how their contributions were valued.

While we analyzed each question separately, we calculated the Spearman rank-order correlation coefficient between each survey question. We found each question’s correlations ranged from weak effect to a moderate effect \(0.0 < \rho < 0.6\). This includes when we sorted by gender, course, and simultaneously gender and course.

IV. DISCUSSION

We compared student outcomes for final course grades and exams based on gender for over 10,000 TAMU students over more than a decade. We examined this data to determine whether such differences were persistent throughout each course. The data were collected from instructors who taught courses from algebra-based or calculus-based introductory physics sequences. Prior studies of gendered students performance based on exam grades show mixed results, with some authors reporting that male students outperform their female counterparts [12, 33, 34], whereas other authors did not find a statistically significant difference between the genders [25, 26, 29, 31, 32, 35, 36]. While this study has the advantage of a large data set of the exam scores collected over a decade and provides a new knowledge on the gendered performance in the introductory physics classes, it has limitations: only course-level data collected from faculty was used in this study. Therefore, we did not analyze the impact of non-academic factors that have seen to po-
To describe our results, we use “significant” as a shorthand for “statistically significant”, defined as $p < 0.05$.

Of the four introductory courses comprising the algebra-based and calculus-based sequences, only the algebra-based mechanics course exhibited a significant difference in final letter grades. This difference is small, with female students outperforming male students by 0.15 GPA points and has a weak effect size. This is in agreement with a prior study by Tai and Sadler who also reported female students performing better than male students in algebra-based mechanics [39]. In algebra-based mechanics, significant performance differences based on gender were found on two out of five exams throughout the course. These were midterm exams in the second half of the course, and the differences had weak effect sizes.

In calculus-based mechanics, there was no significant difference in final letter grades. There was, however, a significant difference on three out of four course exams, with a small effect size for the first exam, and weak effect sizes for the third and fourth exams. Previous studies have found inconsistent results from calculus-based mechanics courses. Some studies conducted at other public, state universities have found no significant gap in final letter grades [26, 29, 30]. Kost et al found a small but statistically significant difference in overall course grades [33]. Also, Tai and Sadler cited above for the algebra-based course, reported males outperforming females in calculus-based mechanics [39].

In the second semester E&M course for both sequences, no significant difference in final letter grades was found. The only significant difference observed was for the first exam in the algebra-based course which had a weak effect size. The results are similar to Kost-Smith et al [32] who also found no statistically significant difference in course grades in a calculus-based electricity and magnetism course.

When examining the impact of instructor gender on the student gender performance gap, fewer significant differences were observed for female instructors in comparison with male instructors. Where significant differences were seen, the effect sizes were small or weak, similar to the effect sizes when data from all instructors was pooled. It should be noted that the overall number of instructors included in this study was small (~20 unique instructors) and a larger sample of instructors would help to determine the consistency of this result.

In addition to analyzing historic exam scores, we took a snapshot of current students perception of their performance as well as their feelings of inclusion and contribution in the introductory physics courses sequences. We administered a questionnaire to all students who took introductory physics classes in the fall of 2019. We received over 1600 completed questionnaires with a response rate of 63%.

The responses indicated that female students rated their performance lower than their male classmates. The one exception is algebra-based mechanics where female students rated their performance equal to male students. Although our question about the performance perception didn’t directly measure students physics self-efficacy, our results are consistent with the previous studies reporting that female students display lower self-efficacy than male students in introductory physics classes, even when controlled for their performance level [12, 16, 21, 29].

We found no significant gendered difference in students’ perceptions of inclusion across all courses. Also, there was no significant differences in students’ perceptions about the value of their contributions, except for algebra-based mechanics, where female students reported that their contributions were valued higher than male students. Gender neutral perception of inclusion and perception about the value of students contributions is positive news taking into account that the questionnaire was distributed towards the end of the semester when students had enough time to form an opinion. In algebra-based courses, the enrollment of students tends to have more female students than male students, and students are typically upper-level undergraduates. As a result, female students in these classes may experience less stereotype threat [11], which could partially explain why they report equal or better perception of inclusion and valuation of their contribution as compared to their male counterparts. In algebra-based mechanics, the instruction during the recitations was provided by upper-level undergraduate students who happened to about 70% females in fall 2019 when the questionnaire was distributed. This could contribute to female students rating the value of their contributions higher than their male classmates [15]. It is also worth noting that female students rated their performance as equal to male students and the value of their contributions higher than male students in the algebra-based mechanics class, where female students historically outperformed male students based on our analyses of more than a decade of exam data.

V. CONCLUSION

This study explored the evolution of gendered differences in student performance throughout two-semester introductory physics courses for two sequences, calculus-based and algebra-based. The performance indicators for a large pool of over 10,000 students spanning a period from spring 2007 to spring 2019 have been analyzed. Data on midterm exams, final exams, and final course grades were collected from instructors teaching these courses during that period of time. Where differences in final letter grades were found, there were no persistent differences on exams. Where persistent differences were found on exams, there were no differences for final letter grades. In algebra-based mechanics, female students outperformed male students by a small but significantly different margin. In all statistically signifi-
cant cases, the effect size was small or weak, indicating that performance on exams and final letter grades was not strongly dependent on gender. One should keep in mind that the instructors who taught these classes used different instruction methods that were evolving during the period of data collected. The student demographics, background, and preparation levels were evolving during the years of study [44]. Since this is a long-term historic study of course-level data, we were not able to control for factors such as student background in math and physics.

A questionnaire was distributed to students taking both calculus-based and algebra-based sequences during the fall 2019 semester. The goal was to take a snapshot of current students’ feelings to see how their perceptions aligned with historic performance. We collected students’ feedback on their perception of their performance, feelings of inclusion, and the value of their contributions. The analyses of student responses revealed no difference in the feeling of inclusion in any of these courses. For one course, algebra-based mechanics, female students rated in the feeling of inclusion as high as their male counterparts. For the other three courses, male students reported higher perceptions of performance than female students.

There are several future studies that could stem from this one. In the next iteration of this work, it would be beneficial to connect course-level data with university records of students prior preparation and knowledge. Additionally, an enhanced survey on student perceptions could be linked with course performance to allow for regression analyses between inclusion and contribution with success on exams. Since calculus-based mechanics is usually taken the earliest of these four courses, it would be useful to perform a study like this one on calculus 1 and introductory chemistry. This would help us better understand if gendered performance differences among physical science and engineering majors change over time.

ACKNOWLEDGMENTS

This study was supported by the Texas A&M University College of Science. We would like to thank the Texas A&M University Department of Physics and Astronomy faculty who provided us with data for this study.

[1] Paula Heron and Laurie McNeil, *Phys21: Preparing Physics Students for 21st-Century Careers* (2016).
[2] Anne Marie Porter and Rachel Ivie, *Women in Physics and Astronomy, 2019* Tech. Rep. (AIP Statistical Research Center, Melville, NY, 2019).
[3] National Science Foundation, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019* Special Report NSF 19-304 (National Center for Science and Engineering Statistics, Alexandria, VA, 2019).
[4] Linda J. Sax, Kathleen J. Lehman, Ramón S. Barthelmeey, and Gloria Lim, “Women in physics: A comparison to science, technology, engineering, and math education over four decades,” *Phys. Rev. Phys. Educ. Res.*, 12, 020108 (2016).
[5] Laura McCullough, “Women in physics: A review,” *The Physics Teacher* 40, 86–91 (2002), https://doi.org/10.1119/1.1457312.
[6] Christine Nord, Shep Roe, Robert Perkins, Marsha Lyons, Nita Lemanski, Yaed Tamir, Janis Brown, Jason Schunkhe, , and Kathleen Herrold, “The nations report card: Americas high school graduates,” (2011), retrieved from https://nces.ed.gov/nationsreportcard/pdf/studies/2011462.pdf.
[7] Irene Goodman, Christine Cunningham, Cathy Lachapelle, Meredith Thompson, Katherine Bittinger, Robert Brennan, and Mario Delsi, “Final report of the women’s experiences in college engineering (wece) project. eric ed507395,” (2002), 10.13140/2.1.2027.5524.
[8] Lin Bian, Sarah-Jane Leslie, and Andrei Cimpian, “Gender stereotypes about intellectual ability emerge early and influence children’s interests,” *Science* 355, 389–391 (2017). https://science.sciencemag.org/content/355/6323/389.full.pdf.
[9] Nancy M. Hewitt and Elaine Seymour, “A long, discouraging climb,” *ASEE Prism* 1, 24–28 (1992).
[10] Gwen C. Marchand and Gita Taasoobshirazi, “Stereotype threat and women’s performance in physics,” *International Journal of Science Education* 35, 3050–3061 (2013) https://doi.org/10.1080/09500693.2012.683461.
[11] Alexandru Maries, Nafis I. Karim, and Chandralekha Singh, “Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics?” *Phys. Rev. Phys. Educ. Res.* 14, 020119 (2018).
[12] Akira Miyake, Lauren E. Kost-Smith, Noah D. Finkelstein, Steven J. Pollock, Geoffrey L. Cohen, and Tiffany A. Ito, “Reducing the gender achievement gap in college science: A classroom study of values affirmation,” *Science* 330, 1234–1237 (2010). https://science.sciencemag.org/content/330/6008/1234.full.pdf.
[13] Vashti Sawtelle, Eric Brewe, and Laird H. Kramer, “Exploring the relationship between self-efficacy and retention in introductory physics.” *Journal of Research in Science Teaching* 49, 1096–1121 (2012). https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.21050.
[14] Zahra Hazari, Gerhard Sonnert, Philip M. Sadler, and Marie-Claire Shanahan, “Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study,” *Journal of Research in Science Teaching* 47, 978–1003 (2010). https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.20363.
[15] Z. Yasemin Kalender, Emily Marshman, Christian D. Schunn, Timothy J. Nokes-Malach, and Chandralekha Singh, “Gendered patterns in the construction of physics
identity from motivational factors,” Phys. Rev. Phys. Educ. Res. 15, 020119 (2019).
[16] Sarah L. Eddy and Sara E. Brownell, “Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines,” Phys. Rev. Phys. Educ. Res. 12, 020106 (2016).
[17] Jennifer Blue, Adrienne Traxler, and Geraldine Cochran, “Resource letter: Gp-1: Gender and physics,” American Journal of Physics 87, 616–626 (2019), https://doi.org/10.1119/1.5114628.
[18] Jayson M. Nissen and Jonathan T. Shemwell, “Gender, experience, and self-efficacy in introductory physics,” Phys. Rev. Phys. Educ. Res. 12, 020105 (2016).
[19] Christine Lindstrom and Manjula Sharma, “Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter?” International Journal of Innovation in Science and Mathematics Education 19 (2010).
[20] Emily Marshal, Zeynep Y. Kalender, Christian Schunn, Timothy Nokes-Malach, and Chandradekha Singh, “A longitudinal analysis of students motivational characteristics in introductory physics courses: Gender differences,” Canadian Journal of Physics 96, 391–405 (2018), https://doi.org/10.1139/cjp-2017-0185.
[21] Tobias Espinosa, Kelly Miller, Ives Aranjo, and Eric Mazur, “Reducing the gender gap in students’ physics self-efficacy in a team- and project-based introductory physics class,” Phys. Rev. Phys. Educ. Res. 15, 010132 (2019).
[22] Rachel Henderson, Vashti Sawtelle, and Jayson Micheal Nissen, “Gender & self-efficacy: A call to physics educators,” The Physics Teacher 58, 345–348 (2020), https://doi.org/10.1119/1.5145533.
[23] Emily M. Marshman, Z. Yasemin Kalender, Timothy Nokes-Malach, Christian Schunn, and Chandradekha Singh, “Female students with A’s have similar physics self-efficacy as male students with C’s in introductory courses: A cause for alarm?” Phys. Rev. Phys. Educ. Res. 14, 020123 (2018).
[24] Adrian Madsen, Sarah B. McKagan, and Eleanor C. Sayre, “Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?” Phys. Rev. ST Phys. Educ. Res. 9, 020121 (2013).
[25] Jennifer Docktor, Kenneth Heller, Charles Henderson, Mel Sabel, and Leon Hsu, “Gender differences in both force concept inventory and introductory physics performance,” AIP Conference Proceedings 1064, 15–18 (2008), https://aip.scitation.org/doi/pdf/10.1063/1.3021243.
[26] Simon Bates, Robyn Donnelly, Cait MacPhee, David Sands, Marion Birch, and Niels R Walet, “Gender differences in conceptual understanding of newtonian mechanics: a uk cross-institution comparison,” European Journal of Physics 34, 421434 (2013).
[27] Adrienne Traxler, Rachel Henderson, John Stewart, Gay Stewart, Alexi Papak, and Rebecca Lindell, “Gender fairness within the force concept inventory,” Phys. Rev. Phys. Educ. Res. 14, 010103 (2018).
[28] Rachel Henderson, Paul Miller, John Stewart, Adrienne Traxler, and Rebecca Lindell, “Item-level gender fairness in the force and motion conceptual evaluation and the conceptual survey of electricity and magnetism,” Phys.
[43] Michael J. Cahill, K. Mairin Hynes, Rebecca Trousil, Lisa A. Brooks, Mark A. McDaniel, Michelle Repice, Jiuqing Zhao, and Regina F. Frey, “Multiyear, multi-instructor evaluation of a large-class interactive-engagement curriculum,” Phys. Rev. ST Phys. Educ. Res. 10, 020101 (2014).

[44] D. Martin, Student Enrollment Profile Report. Enrollment Profile (Texas A&M University, 1244 TAMU, College Station, Texas 77843-1244, 2017).

[45] Casper Strongren, “Genderizeio,” (2013), available at https://genderize.io/.

[46] Morten W Fagerland, “t-tests, non-parametric tests, and large studies paradox of statistical practice?” BMC medical research methodology 12, 78 (2012).

[47] Louis Cohen, Lawrence Manion, and Keith Morrison, Research Methods in Education (Routledge, New York City, New York, 2018).
TABLE VII. Average z-score differences between male and female students, \( \Delta \), separated by instructor gender, as well as the t-tests and significances between these distributions.

|                | Male Instructors |               | Female Instructors |               |
|----------------|------------------|---------------|-------------------|---------------|
|                | \( \Delta \)    | t-test        | \( p \)           | \( \Delta \)  | t-test        | \( p \)           |
| Mech. Alg.     |                  |               |                   |               |               |
| Exam 1         | 0.030            | 0.484         | 0.628             | -0.018        | -0.126        | 0.900             |
| Exam 2         | 0.020            | 0.331         | 0.741             | -0.187        | -1.291        | 0.198             |
| Exam 3         | -0.118           | -1.919        | 0.055             | -0.178        | -1.229        | 0.221             |
| Exam 4         | -0.187           | -3.053        | 0.002             | -0.008        | -0.056        | 0.955             |
| Exam 5         | -0.047           | -0.758        | 0.448             | -0.083        | -0.575        | 0.566             |
| Mech. Calc.    |                  |               |                   |               |               |
| Exam 1         | 0.242            | 6.921         | <0.001            | 0.233         | 3.554         | <0.001            |
| Exam 2         | 0.019            | 0.532         | 0.595             | 0.054         | 0.826         | 0.409             |
| Exam 3         | 0.084            | 2.387         | 0.017             | -0.001        | -0.010        | 0.992             |
| Exam 4         | 0.115            | 3.260         | 0.001             | 0.060         | 0.912         | 0.362             |
| E&M Calc.      |                  |               |                   |               |               |
| Exam 1         | 0.128            | 1.404         | 0.161             | 0.041         | 0.731         | 0.465             |
| Exam 2         | 0.054            | 0.595         | 0.552             | 0.037         | 0.647         | 0.518             |
| Exam 3         | 0.034            | 0.370         | 0.711             | 0.021         | 0.362         | 0.717             |
| Exam 4         | 0.307            | 3.384         | 0.001             | -0.020        | -0.352        | 0.725             |

TABLE VIII. Average final letter grade difference between male and female students, \( \Delta \), separated by instructor gender, as well as the t-tests and significances between these distributions.

|                | Male Instructors |               | Female Instructors |               |
|----------------|------------------|---------------|-------------------|---------------|
|                | \( \Delta \)    | t-test        | \( p \)           | \( \Delta \)  | t-test        | \( p \)           |
| Mech. Alg.     |                  |               |                   |               |               |
| Exam 1         | -0.146           | -2.255        | 0.024             | -0.175        | -1.450        | 0.149             |
| Exam 2         | 0.012            | 0.282         | 0.778             | 0.031         | 0.449         | 0.654             |
| Exam 3         | 0.141            | 1.351         | 0.177             | -0.167        | -2.693        | 0.007             |

TABLE IX. Average student performance perception (and standard error) for each course, as well as the t-test and significance between these distributions.

|                | Male Instructors | Female Instructors | t-test | \( p \) |
|----------------|------------------|-------------------|--------|--------|
| Mech. Alg.     | 3.384 (0.081)    | 3.384 (0.062)    | -0.003 | 0.998  |
| E&M Alg.       | 3.342 (0.141)    | 2.886 (0.155)    | 2.154  | 0.035  |
| Mech. Calc.    | 3.395 (0.048)    | 3.062 (0.073)    | 3.699  | <0.001 |
| E&M Calc.      | 3.545 (0.038)    | 3.391 (0.069)    | 2.000  | 0.046  |

TABLE X. Average student feelings of inclusion (and standard error) for each course, as well as the t-test and significance between these distributions.

|                | Male Instructors | Female Instructors | t-test | \( p \) |
|----------------|------------------|-------------------|--------|--------|
| Mech. Alg.     | 3.767 (0.111)    | 3.813 (0.093)    | -0.318 | 0.751  |
| E&M Alg.       | 4.026 (0.188)    | 4.229 (0.146)    | -0.828 | 0.410  |
| Mech. Calc.    | 3.994 (0.063)    | 4.031 (0.091)    | -0.316 | 0.752  |
| E&M Calc.      | 3.964 (0.049)    | 4.005 (0.080)    | -0.431 | 0.667  |
TABLE XI. Average student contribution valuation (and standard error) for each course, as well as the t-test and significance between these distributions.

| Course     | Male       | Female     | t-test | p   |
|------------|------------|------------|--------|-----|
| Mech. Alg. | 3.555 (0.079) | 3.793 (0.056) | -2.52  | 0.012 |
| E&M Alg.   | 3.474 (0.165) | 3.343 (0.189) | 0.515  | 0.608 |
| Mech. Calc.| 3.727 (0.049) | 3.700 (0.078) | 0.289  | 0.773 |
| E&M Calc.  | 3.803 (0.038) | 3.769 (0.067) | 0.453  | 0.650 |