Comparing introductory undergraduate physics learning and behavior before and after the COVID-19 pandemic

Amanda Nemeth, Christopher Wheatley, and John Stewart*

Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA

(Received 2 March 2023; accepted 26 April 2023; published 24 May 2023)

This study examines high school preparation measures [ACT/SAT scores, high school grade point average (HSGPA), and conceptual physics pretest scores], in-class behavior measures (homework submission rates and lecture attendance rates), and in-class achievement measures (homework and test averages) for the last two fully face-to-face prepandemic and the first two fully face-to-face postpandemic semesters of an introductory calculus-based electricity and magnetism class. This class was offered at a large eastern land grant university in the United States. The total number of students for the four semesters was 1033. While some significant differences were measured (higher postpandemic HSGPA, lower postpandemic conceptual pretest scores, higher postpandemic homework average (fall semesters only), and lower postpandemic lecture attendance (spring semesters only), none were larger than a small effect. As such, student achievement, attendance rates, and assignment completion rates were largely unchanged after the pandemic.

DOI: 10.1103/PhysRevPhysEducRes.19.013103

I. INTRODUCTION

Many studies have explored the effects of the COVID-19 pandemic and the rapid transition to virtual modes of education on student attitudes and learning during this period of online instruction [1–3]. Now that most universities in the United States have returned to in-person classes, the effects of this period of disruption on the return to in-person classes can be measured. This study explores differences in student behavior and achievement between the last two fully face-to-face semesters of a college physics class prior to the pandemic and the first two fully face-to-face semesters after the pandemic.

A recent study at a highly selective West Coast University in the United States found that there was no evidence for a reduction in high-school physics learning after the pandemic using a physics diagnostic exam administered in the Fall 2019 and Fall 2021 semesters [4]. This institution is situated in a state with one of the highest per capita incomes in the United States and with a high rate of residents with a bachelor's degree. The current study examines a broader collection of student achievement and behavior measures at an institution admitting students with lower levels of high school achievement than the West Coast University. This institution accepts 90% of its applicants. The institution is the flagship state university in a small eastern state with a state population that ranks among the lowest in the United States in regard to per capita income and rate of bachelor's degrees. As such, it can provide context for the effect of the pandemic on the education of students coming from less-resourced school systems.

II. METHODS

A. Sample

Data were collected from the introductory calculus-based electricity and magnetism course at a large eastern U.S. land-grant university with a total undergraduate enrollment in Fall 2021 of 19 600 students [5]. The demographics of the undergraduate population were 81% White, 4% Black or African American, 4% Hispanic or Latino, 3% nonresident alien, 6% two or more races, with other groups 2% or less. Student ACT scores ranged from 21 to 27 for the 25th to the 75th percentile. Pell grants are given to lower socioeconomic status students (SES) and are often used to measure the percentage of lower SES students at a university: 23% of the undergraduate population was Pell eligible. The Fall 2021 enrollment was smaller than that of Fall 2019 when the university enrolled 21 000 students. The overall demographics of the Fall 2019 undergraduate population were 80% White, 4% Black or African American, 4% Hispanic/Latino, 5% nonresident alien, 4% two or more races, with other groups 2% or less.

The course studied enrolled primarily scientists and engineers. The course was taught in multiple lecture sections that were overseen by the same lead instructor for the entire period studied. This instructor had been managing the course for many years and oversaw general

*jcstewart1@mail.wvu.edu

course content, homework assignments, tests, and the management of the laboratory segment of the class. The class had been offered in the same format for many years before the pandemic and returned to this format after the pandemic. Alternate modes of instruction were provided during the pandemic; these semesters are not considered in this study. As such the course represents an excellent laboratory to study changes before and after the pandemic. This study focuses on the introductory electricity and magnetism course because the introductory mechanics course retained some pandemic course policy changes preventing comparison.

The class was presented with three 50-min lecture sessions along with one 170-min laboratory session per week. Each lecture session enrolled over 100 students. Both the lecture and laboratory utilized multiple active learning strategies. The lectures implemented Peer Instruction using clickers [6], while the labs featured a mixture of conceptual whiteboard questions, hands-on inquiry activities, group problems, and traditional experiments. Two homework sets were collected each week; these were collected at the beginning of the lecture and were turned in on paper. The first homework collected on Monday of each week, called the "short homework" in this study, consisted of ten multiple-choice questions. The second homework collected on Wednesday each week, called the "long homework" in this study, consisted of five multiple-choice questions and four open-response questions. Four tests and a final exam were given over the course of the semester. Conceptual learning was monitored by applying the Conceptual Survey of Electricity and Magnetism (CSEM) [7] as a pretest and post-test. The class was primarily taken by sophomores, who would have spent much of their freshman year in college and the last part of their senior year in high school receiving virtual instruction. These students likely took the ACT or SAT and submitted a college application prior to the pandemic.

The total enrollment for the four semesters was $N = 1033$ (Spring 2019 $N = 217$, Fall 2019 $N = 327$, Fall 2021 $N = 327$, Spring 2022 $N = 162$). Assignment scores and lecture attendance were accessed from the course learning management system; students who withdraw from the class are automatically removed from course records. As such, the analysis, except the DFW percentage, includes only students who completed the class for a grade.

B. Statistical methods

Often in PER studies, t tests are used to compare means; however, the t test assumes the sample is normally distributed. Many of the quantities examined in this work (homework scores, lecture attendance rates, etc.) have distributions that are substantially non-normal. As such, this work applies the nonparametric Mann-Whitney U test. The Mann-Whitney U test, sometimes referred to as the Wilcoxon rank-sum test, is a rank-sum test used to determine whether the total ranks of two independent groups significantly differ. The test calculates U which is related to the sum of the ranks for one of the groups. An effect size, r, can be calculated from U; Cohen's criteria for r are that 0.10 is a small effect, 0.30 a medium effect, and 0.50 a large effect [8,9].

Some quantities examined were semester-level frequencies such as the DFW percentage (percentage of students earning a grade of D or F or withdrawing from the class); these were compared using the two proportions z test. The two proportions z test, which is equivalent to a chi-squared test for the equality of two proportions, is a test for checking if the difference between two proportions is statistically significant. The corresponding effect size is Cohen's h; the effect size criteria for h are similar to Cohen's d, 0.2 corresponds to a small effect, 0.5 to a medium effect, and 0.8 to a large effect. Both the Mann-Whitney U test and the two proportions z test are discussed in more detail in the Supplemental Material [10].

This work applies many statistical tests and is, therefore, susceptible to the inflation of type I error. The Holm-Bonferroni correction for the significance level is applied [11]. This method orders the p values from smallest to largest, then progressively adjusts the significance level. If there are m statistical tests, the significance threshold, α, is adjusted to α/m for the smallest p value, $\alpha/(m-1)$ for the second smallest, etc. The null hypothesis is rejected for all p greater than the first p which fails the test. This method provides the same type I error correction as the Bonferroni correction with less risk of type II error.

III. RESULTS

Assignment scores and submission rates were compared between the last two completed semesters preceding the COVID-19 pandemic shutdowns (Spring and Fall 2019) and the first two semesters after in-person courses were resumed (Fall 2021 and Spring 2022). The institution did not return to remote instruction at any time in the post-pandemic period. The assignment submission rate is the percentage of the assignments submitted for grading. The assignment percentage score is the average score on the assignment (zero if not submitted). The two fall semesters were compared against each other pre- and postpandemic, as were the two spring semesters. The class studied has historically observed a substantial difference in student performance between the spring and fall semesters which is likely the result of differences in the high school preparation of the students in these semesters. For the class studied, students in the fall semester are "on-sequence" pursuing the plan of study suggested by the university's 4-year degree plans; these students were ready to enroll in calculus 1 in their first semester in college; largely students in the spring semester were not.

A summary of general descriptive statistics is shown in Table I. The results are reported by semester with Fall 2019

TABLE I. Mean ± standard deviation by semester. Pairs of fall and spring semesters are compared with a Mann-Whitney U test, the U statistic; its p value, z score, and effect size r are also reported. Bolded p values are significant at the $p < 0.05$ level after a Holm-Bonferroni correction is applied. The p value reported is the uncorrected value.

Variable	Semester	N	M ± SD	U	p	z	r
Test average	F19	307	76.1 ± 14	49217	0.188	1.32	0.05
	F21	302	74.2 ± 15				
	S19	206	71.6 ± 16	15472	0.645	0.46	0.02
	S22	146	70.3 ± 18				
High-school GPA	F19	298	3.89 ± 0.45	33781	**0.000**	−4.94	0.20
	F21	296	4.07 ± 0.39				
	S19	198	3.76 ± 0.46	10008	**0.000**	−3.99	0.22
	S22	136	3.96 ± 0.43				
Lecture attendance percentage	F19	307	82.6 ± 23	42690	0.082	−1.74	0.07
	F21	302	86.0 ± 21				
	S19	206	84.5 ± 25	17736	**0.003**	2.97	0.16
	S22	146	78.0 ± 27				
ACT/SAT mathematics percentile score	F19	285	84.9 ± 15	45610	0.046	2.00	0.08
	F21	292	85.2 ± 11				
	S19	183	79.9 ± 17	11486	0.855	0.18	0.01
	S22	124	79.0 ± 18				
Short homework percentage score	F19	307	67.6 ± 24	42458	0.072	−1.80	0.07
	F21	302	71.3 ± 22				
	S19	206	67.7 ± 25	15165	0.893	0.14	0.01
	S22	146	68.1 ± 25				
Long homework percentage score	F19	307	62.4 ± 23	38187	**0.000**	−3.76	0.15
	F21	302	68.8 ± 22				
	S19	206	60.0 ± 25	13610	0.129	−1.52	0.08
	S22	146	63.5 ± 25				
Short homework submission percentage	F19	307	83.8 ± 23	41773	0.025	−2.25	0.09
	F21	302	86.6 ± 22				
	S19	206	84.0 ± 24	16376	0.132	1.51	0.08
	S22	146	80.6 ± 27				
Long homework submission percentage	F19	307	82.7 ± 23	41611	0.020	−2.33	0.09
	F21	302	87.0 ± 22				
	S19	206	83.9 ± 25	16409	0.117	1.57	0.08
	S22	146	79.8 ± 28				
CSEM pretest percentage	F19	298	27.7 ± 10	49027	**0.002**	3.08	0.13
	F21	287	25.7 ± 11				
	S19	192	27.0 ± 12	13548	0.483	0.70	0.04
	S22	135	26.0 ± 11				
CSEM post-test percentage	F19	247	59.4 ± 17	32266	0.042	2.04	0.09
	F21	236	56.4 ± 17				
	S19	184	59.2 ± 18	11708	0.016	2.40	0.14
	S22	109	54.0 ± 18				

abbreviated F19. Most quantities in Table I did not significantly change between pre- and postpandemic semesters after applying the Holm-Bonferroni correction. The p values are reported without correction in the table and bolded if they meet the adjusted significance threshold. For both semesters, high-school GPA (HSGPA) was significantly higher, a small effect, in postpandemic semesters. This may have been a result of the changes imposed on high school instruction and grading by the pandemic. While lecture attendance in the fall semesters did not

TABLE II. Difference in means between course-level variables where *M%* is the percentage of students reporting ACT/SAT scores or the percentage of DFW students, *p* is the *p* value comparing semesters, and *h* is the effect size of the difference.

Variable	Semester	*N*	*M%*	χ^2	*p*	*h*
ACT/SAT reporting percentage	F19	274	97.5	0.48	0.490	0.08
	F21	291	98.6			
	S19	178	97.3	3.90	0.048	0.25
	S22	121	91.7			
DFW percentage	F19	47	14.4	0.42	0.516	0.06
	F21	54	16.5			
	S19	42	19.4	0.49	0.485	0.09
	S22	37	22.8			

change through the pandemic, lecture attendance in the postpandemic spring semesters was significantly lower, also a small effect. Although HSGPA increased after the pandemic, CSEM pretest scores significantly decreased in the fall semesters. This effect is functionally negligible, representing less than one additional pretest question answered correctly before the pandemic.

Table II shows the DFW percentage—the percentage of students who received a D in the class, an F in the class, or withdrew from the class—and the ACT/SAT score reporting percentage for the class. Only domestic students are included in the ACT/SAT results. International students are less likely to submit ACT/SAT scores than U.S. students. International attendance dropped in this class from 10.4% in Fall 2019 to 3.3% in Fall 2021. Neither the DFW percentage (all students) nor the ACT/SAT reporting percentage (domestic students) were significantly different pre- and postpandemic.

To understand how student behavior changed over the course of a semester, this study looked at the submission percentage of homework disaggregated by the four in-semester examinations testing the material covered by the homework. The long homework submission percentages are shown in Fig. 1; the short homework percentages, which are very similar, are shown in the Supplemental Material [10]. All homework assignments which were due during the part of the course covered by each examination were included in the average for the test. For both types of homework, the rate at which students submitted homework assignments was higher in the postpandemic Fall 2021

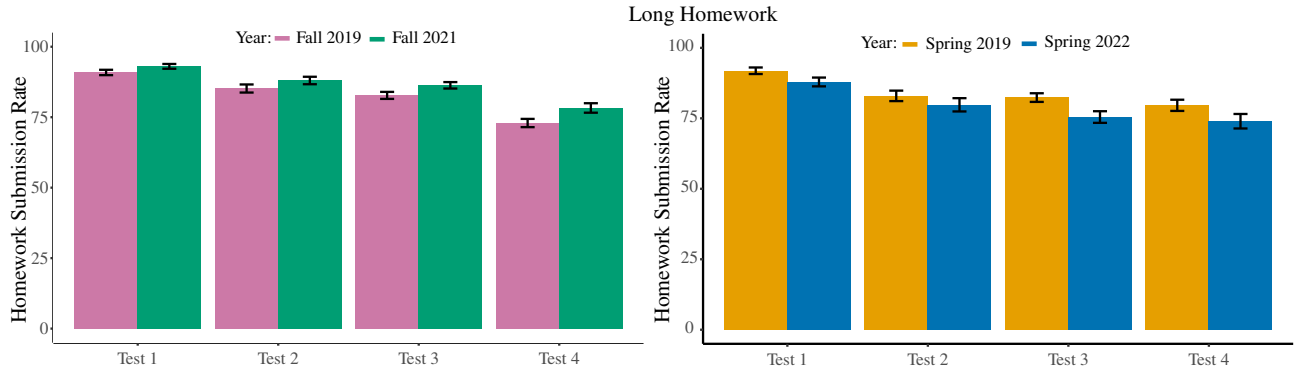

FIG. 1. The average submission rates for the long homework. The rate is the percentage of the homework assignments submitted for grading.

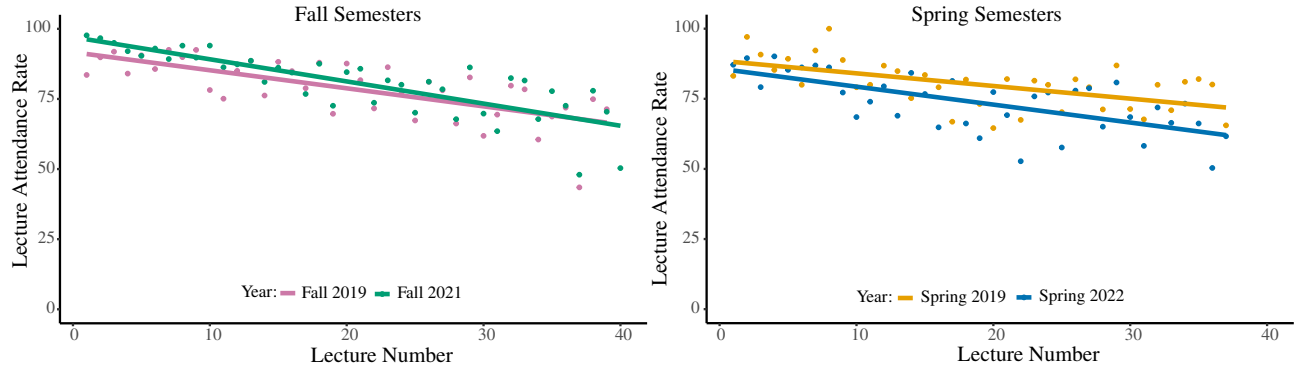

FIG. 2. Average lecture attendance rates plotted against the order in which the lecture was given. The rate is the percentage of students attending each lecture section.

semester for every test, while for both types of homework that rate was lower in the postpandemic Spring 2022 semester for all tests. These differences were, however, small; none of the differences were statistically significant after applying the Holm-Bonferroni correction.

This study also examined the evolution of lecture attendance over the semester (Fig. 2). In these plots, lecture number represents the order of the lecture in the semester and is, therefore, a rough measure of time. For the fall semesters, students attended lectures at a higher rate in Fall 2021 early in the semester, but the rates equalized late in the semester. For the spring semesters, students attended lectures at a lower rate in Spring 2022 throughout the semester and the difference in rates became larger later in the semester. In both cases, the rate of lecture attendance decreased at a larger rate in the postpandemic semesters (simple linear regression slopes predicting submission rate with lecture number as an independent variable: Fall 2019:−0.013, Fall 2021: −0.016; Spring 2019: −0.009, Spring 2022: −0.013). Regression analysis demonstrated these slopes were significantly different ($p < 0.001$) with significant spring or fall effects and significant pre- to postpandemic effects. These differences caused the attendance rates for the fall semesters to converge at the end of the semester while the attendance rates for the spring semesters diverged.

IV. DISCUSSION AND CONCLUSION

The course studied is typically taken by sophomores; most students enrolled in the course are "on-sequence" if they enroll in the course in their fall sophomore semester. Those students likely had a pandemic-interrupted high school experience in their final senior semester. They would have taken their ACT or SAT tests prior to the pandemic. High school physics is often taken in the senior year; as such, the transition to online instruction may have affected their high school physics class; this would explain differences in CSEM pretest scores. Most students in the class would have taken both calculus 1 and the introductory mechanics physics class as fully online classes.

In general, few significant differences were measured between pre- and postpandemic behavior and academic achievement; all significant differences were small effects. This suggests the results of Burkholder and Wieman [4] showing students' physics preparation at a highly selective U.S. institution was unchanged through the pandemic extend to students at less selective institutions and are fairly general across achievement on physics assignments and rates of turning in assignments and attending classes.

The few significant differences measured were consistent with the qualitative impression of course personnel but smaller than they expected. Fall 2021 marked the first semester back to primarily in-person instruction, the course personnel reported that students were generally enthusiastic to return to in-person instruction. This enthusiasm might also explain the statistically significant increase found in long homework scores in Fall 2021. The initial higher lecture attendance in the fall semester postpandemic decreased over the course of the semester until the fall lecture attendance became equal pre- and postpandemic. This is consistent with an initial enthusiasm for a return to face-to-face instruction which declined over the semester. Course personnel also felt that student engagement was lower and declined over the semester in the Spring 2022, particularly after spring break. This is also supported by the growing gap between attendance rates for spring semesters pre- and postpandemic. Course personnel were expecting larger differences from this study, possibly because they were anticipating a substantial effect of the pandemic on student performance and behavior. The actual differences observed were quite small.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grants No. ECR-1561517, No. HRD-1834569, and No. DUE-1833694 and by a grant from the Howard Hughes Medical Institute.

[1] P. Klein, L. Ivanjek, M. N. Dahlkemper, K. Jeličić, M.-A. Geyer, S. Küchemann, and A. Susac, Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories, Phys. Rev. Phys. Educ. Res. **17**, 010117 (2021).

[2] M. F. J. Fox, J. R. Hoehn, A. Werth, and H. J. Lewandowski, Lab instruction during the COVID-19 pandemic: Effects on student views about experimental physics in comparison with previous years, Phys. Rev. Phys. Educ. Res. **17**, 010148 (2021).

[3] I. Marzoli, A. Colantonio, C. Fazio, M. Giliberti, U. Scotti di Uccio, and I. Testa, Effects of emergency remote instruction during the COVID-19 pandemic on university physics students in Italy, Phys. Rev. Phys. Educ. Res. **17**, 020130 (2021).

[4] E. W. Burkholder and C. E. Wieman, Absence of a COVID-induced academic drop in high-school physics learning, Phys. Rev. Phys. Educ. Res. **18**, 023102 (2022).

[5] National Center for Education Statistics, https://nces.ed.gov/collegenavigator.

[6] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
[7] D. P. Maloney, T. L. O'Kuma, C. Hieggelke, and A. Van Huevelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).
[8] G. W. Corder and D. I. Foreman, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach* (John Wiley & Sons, Inc., Hoboken, NJ, 2009), pp. 57–72.
[9] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, NY, 1977).
[10] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.19.013103 for a thorough description of the statistics used in the paper and the plot of the submission rate of the short homework.
[11] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. **6**, 65 (1979).