Check for updates

Received: 18 May 2021 | Accepted: 23 March 2023

DOI: 10.1002/soej.12635

RESEARCH ARTICLE

Southern Economic Journal WILEY

In-person versus online instruction: Evidence from principles of economics

Kenneth G. Elzinga | **Daniel Q. Harper**

Department of Economics, University of Virginia, Charlottesville, Virginia, USA

Correspondence
Kenneth G. Elzinga, Department of Economics, University of Virginia, PO Box 400182, Charlottesville, Virginia 22904, USA.
Email: elzinga@virginia.edu

Abstract

COVID-19 required many professors to switch from in-person teaching to online instruction, allowing exploration of a pivotal question in education: are learning outcomes better when instruction takes place in-person or online? We compare student performance across two semesters of the same large introductory economics course—one taught in-person in 2019, the other taught online in 2020. We analyze test scores from over 2000 students for exam questions common to both instructional formats. At the aggregate level, we find no difference in student performance between online and in-person instruction. When dividing questions by required reasoning skills, we find that online instruction improves student performance on questions requiring knowledge of a definition or formula. Additionally, student course evaluations rated the online course over in-person pedagogy.

JEL CLASSIFICATION

A22

1 | INTRODUCTION

Along with its challenges, COVID-19 brought an opportunity: learning whether students perform better when instruction takes place live and in-person or remotely and online. Put differently, in terms of students learning the material, does it matter whether classes meet face-to-face in real time, or meet digitally and asynchronously?

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Southern Economic Journal* published by Wiley Periodicals LLC on behalf of The Southern Economic Association.

A great deal is at stake in the answer to this question. If student educational outcomes are the same whether material is taught in a classroom or through a device, the primary economic question is: which pedagogical format is less costly to supply? However, if student educational outcomes are better in one teaching format relative to the other, the economic question becomes: how much better, relative to any cost differences between the two formats?

Comparisons of in-person versus online learning have been explored before. Some studies find a negligible difference between online learning and in-person learning (Gratton-Lavoie & Stanley, 2009). Sosin et al. (2004) find positive effects from online instruction. Other scholars conclude that online instruction has a negative effect on performance (Brown & Carl, 2002; Calafiore & Damianov, 2011; Farinella, 2007). Research on this topic is summarized in Miller and Rebelein (2012). These early studies rely on data gathered from courses where students endogenously choose to take an in-person or online class.

More recent studies have randomly assigned students to in-person or online classes. Figlio et al. (2013) performed one of the first randomized assessments of live versus Internet instruction. Like our paper, the students were drawn from large principles of microeconomics course at a research university. Unlike our study, students were randomly assigned to live versus remote lectures, after endogenously opting into the study. As with our study, other variables were held constant (such as text and exams). As measured by test scores, the authors found that live instruction produced slightly higher educational attainment for Hispanic, male and "lower-achieving students." While their study limits endogeneity concerns by randomly assigning students to online and in-person lectures, our natural experiment also removes the endogeneity of students opting into the study.

Alpert et al. (2016) also used students in a principles of economics course to assess learning outcomes of online education. The random design of their paper considered three pedagogical formats: live, online, and a blend of the two. The authors found that students who completed the online course had test scores below those in the live class. This was not the case for students in the blended class, suggesting that colleges might be able to reduce instructional costs without impeding learning by offering part of a principles course online. While we have a similarly sized data set as their study, ours is gathered over only two semesters with all students taking the class with the same instructor. Additionally, we only compare differences between online and in-person instruction.

In a recent paper, Kofoed et al. (2021) evaluated student performance after the COVID-19 pandemic mandated online instruction at West Point. Students were cadets who, unlike students in our study, were taught in small classes by 12 different instructors. After examining both assignments and test results, the authors found that—compared to in-person teaching—online education had negative results, with the adverse consequences being greatest for academically at-risk students.

Scholars comparing educational performance usually encounter the same problem: evaluating different pedagogical formats under ceteris paribus conditions. If researchers use an across-semester approach, comparing some metric of student performance across different semesters of a course, it is difficult to hold constant the other variables that might affect student performance. If a cross-section approach is used, researchers must somehow ensure that the only difference between the two subject groups is that some students were taught in-person, others online.[1] In this paper, we take the across-semester approach, focusing on the Fall 2019 and

[1]When the COVID-19 pandemic first affected colleges and universities, it appeared that there might be many opportunities to examine student outcomes across in-person and online learning for students at the same institution, taking the same course, at the same level of difficulty. However, as the severity of the pandemic became better known, fewer courses were offered in-person, and more were offered online. As a non-random example, during the Fall 2020 semester, the Department of Economics at the University of Virginia offered 43 undergraduate courses and sections of courses. Of these, only seven were offered in-person. In the case of these seven, the faculty were obligated to offer the courses online as well.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2020 semesters. Fall 2019 was the last full semester before COVID-19 and Fall 2020 was the first full semester during COVID-19. We explain in the following section how confounding variables other than the method of instruction are held constant.

Generally, we find no overall difference in student performance between online and in-person instruction. We do find evidence that out-of-state students had lower performance from online instruction during COVID-19. Additionally, we find that when dividing questions by different cognitive skills, online learning results in improved performance from questions that require the recall of definitions or manipulation of formulas. These results contrast with Orlov et al. (2020), who find that online instruction during COVID-19 had an overall negative impact on student performance. However, their study focuses on the Spring 2020 semester, when instructional methods abruptly changed mid-semester. In our study, the instructional method is constant throughout each semester.

2 | THE FRAMEWORK

2.1 | Overview of course and administration

The COVID-19 pandemic and the large number of students enrolled in the Principles of Microeconomics course at the University of Virginia (hereafter Econ 2010) offered a unique opportunity to evaluate the merits of in-person versus online instruction.[2] Like many introductory economics courses across the country, enrollment in Econ 2010 tilts toward first- and second-year students. These students are testing the waters to help them choose their major field of study. Even though most of the students in Econ 2010 are first- and second-years, the course also attracts many third- and fourth-year students who, while majoring in other subjects, want to experience a course in economics before graduating. While many of the students in the course major in economics, business, or public policy (degrees for which the course is required), the course also counts toward the Social and Economic Systems general education requirement. About 1/3 of all undergraduate students at the University of Virginia take Econ 2010, and 1/4 of all undergraduate students take Econ 2010 with Elzinga. This study uses data from the Econ 2010 courses taught by Elzinga.[3]

One feature that increases student demand for Econ 2010 at the University of Virginia is the presence of two other undergraduate schools: the McIntire School of Commerce and the Batten School of Leadership and Public Policy. Entry to these schools is selective and open only to third- and fourth-year students. Econ 2010 is a prerequisite for admission to both.[4]

In the Fall of 2019, Elzinga's Econ 2010 course was taught in-person (as it had been for years) in a large lecture hall. The discussion sections also were taught in-person. However, in

[2]In an earlier paper, Elzinga and Melaugh (2009) used data from multiple years of this course to evaluate the relative performance of students in introductory economics based on their year of enrollment, school of enrollment, gender, and race.

[3]In Fall 2019 other sections of Econ 2010 were offered in addition to Elzinga's sections. In Fall 2020, Elzinga's Econ 2010 course was the only one offered.

[4]While Econ 2010 is not considered a "weed out" course for the two schools (in the manner that Organic Chemistry at many schools is considered a "weed out" course for pre-med students), students aspiring for entry to both the Batten School and the McIntire school understand that a low grade in Econ 2010 reduces their admission prospects. In 2020, the Batten School of Leadership and Public Policy admitted 92 out of 216 applicants (a 42% acceptance rate). The Commerce School admitted 358 out of 568 applicants (a 63% acceptance rate).

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the Fall of 2020, during the COVID-19 pandemic, the course and all discussion sections were taught online as mandated by the University's administration. The lectures were presented asynchronously while the discussion sections and exams took place synchronously.

Several important variables that could affect student performance were held constant across the 2 years. The textbook, David Colander's *Microeconomics* (11e), was used both years. The reading assignments from this textbook were the same. The test format for both years was the same: two midterm exams and one final exam. The person giving the lectures (Elzinga) was the same and the Head Teaching Assistant (Harper) was the same. Course credit hours (3) also were unchanged.

In addition to the course instructor and Head Teaching Assistant remaining constant, the composition and quality of the Teaching Assistants were comparable across both years. In 2019, 73.9% of the Teaching Assistants were undergraduates; the rest were graduate students in the Department of Economics. In 2020, 71.4% of the Teaching Assistants were undergraduates; the rest were graduate students.[5] In 2020, 21.4% of the Teaching Assistants in 2020 had been Teaching Assistants in 2019. During both years the Teaching Assistants had the same responsibility of leading their respective discussion sections, holding office hours, and grading two midterm exams and a final exam. The only major change for Teaching Assistants in 2020 relative to 2019 was that they taught their discussion sections and held office hours remotely, usually by Zoom, rather than in-person as in 2019.

The student population also was constant. In 2019, 60.5% of students were first-years; in 2020, 63.5% were first-years. While the student body may grow slightly more competitive every year, students taking Econ 2010 in 2019 and 2020 are a near identical sample.[6]

Apart from the method of delivery, the content of the lectures in Econ 2010 was comparable across the two semesters. The lectures were tweaked for the Fall of 2020 only to improve exposition or to make the material timelier. For example, charts and numbers were updated where appropriate. However, the topics covered were identical across both years and would be considered standard fare in principles of microeconomics class.

The main difference in the lectures was the format. Normally, Econ 2010 is taught in a large auditorium, in two back-to-back sections of approximately 500 students each. This was the case in 2019. In 2020, however, the lectures were pre-recorded during the summer in a soundproof video studio, using two cameras, two mics, and studio lighting. In both cases, the lectures were presented with accompanying PowerPoint slides.

An ancillary difference in the presentation of the lectures between the 2 years was their time continuity and availability: the in-person lectures in 2019 were continuous for the entire 55-min class period. These 55-min lectures were only available once to students who attended class. In 2020, the lectures were divided into segments of 10–30 min, usually three to four per lecture. This meant students could take a break while listening to a lecture, a feature that was not an option during the live classroom presentation. In the Fall of 2020, students had access to the lecture recordings for 72 h after they were made available online. After 72 h, access was restricted.

[5]Selection as an undergraduate Teaching Assistant is competitive. Virtually all undergraduate TAs received at least an A in the principles course and are in the top 10% of their class. They are chosen based on their grades and an application essay. At the beginning of the semester all TAs attend an orientation meeting, where TA responsibilities are discussed. Many departments at UVA (including the Economics Department) favor a model of instruction where classes at all levels have a TA and discussion section. The upper year students selected as TAs are therefore familiar with what is expected in a discussion section at UVA. Graduate TAs are assigned by the Director of Graduate Studies. Graduate TAs attend a one-day workshop on teaching before their first TA assignment.

[6]In 2019, the acceptance rate at UVA was 23.94%. In 2020, the acceptance rate was 22.58%.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This was done to encourage regular and timely class attendance and discourage binge-watching lectures for the first time prior to a test. Students could watch these videos as many times as they desired in the 72-h period. The lectures were made available again for 5 days prior to the final exam for review. In the Fall of 2019, students did not have recordings of the lectures to use as study material before the final exam.

Offering the course remotely meant the class became scalable. Enrollment was no longer limited to two sections of 500 students (1000 total students), which was the capacity of the largest teaching auditorium at the University of Virginia. In 2020, all students who wanted to enroll could be accommodated. The number of students increased to approximately 1250 students.[7]

2.2 | Assessments and grading

For purposes of our study, a key variable that was held constant was the composition and grading of the final exam. Because the final exam was multiple choice, there was no variability in grading across Teaching Assistants or years.

The total course grade was based on the final exam, two midterm exams, performance on in-class clicker questions, and a discussion section grade. These assessments are individually scored out of different numbers of points to sum to a total of 400 points for the course grade. In 2019 the weighting of these assessments was final exam (200 points), midterm exams (80 points each), clicker questions (15 points), and discussion section grades (25 points). The move to online instruction made it impossible to implement in-class clicker questions the same as in 2019. These were replaced by after-lecture quizzes that were due within 24 h of a set of lecture videos being posted, which were longer and involved more questions than the in-class clicker questions. To adjust for this difference the 2020 weighting was final exam (180 points), midterm exams (80 points each), lecture quizzes (30 points), and discussion section grades (30 points).

Discussion section grades were awarded by a student's Teaching Assistant. TAs had autonomy over how they chose to award these points, with the caveat that no more than five points could be awarded solely for attendance. Additionally, TAs were instructed to aim for an average of 80% for their discussion section points. Popular options included homework assignments and in-class quizzes. The method of assessment for these points varied significantly by TA, making discussion section points an undesirable margin of comparison for online versus in-person instruction.

Midterm exams were free-response style tests graded by a student's TA using a common key. While the weight of the midterm exams was consistent across the two methods of instruction, the move to online learning resulted in changes to the style of questions. This was most notable with questions that required students to use a graph to answer a question. In 2019, such questions accounted for 37.5% of the points on midterm 1 and 25% on midterm 2. With online instruction in 2020, the testing software did not allow for graphing questions of the style used in the in-person course. To accommodate this, questions that previously required graphs were rewritten to require a combination of fill in the blank and short answer responses. This change makes a cross-year analysis of the midterm exam scores undesirable.

[7]In 2020, 1290 seats for Econ 2010 were available, with not all seats filled. In 2019, there were 990 seats available with Elzinga and 225 seats available in a section taught by another instructor, for a total of 1215 seats available. Thus, in net there were 75 additional seats in Econ 2010 offered in 2020.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Between the 2019 final exam and the 2020 final exam, 61 out of 75 questions were identical. All of these common questions were written by the instructor, no questions were from a test bank. Because of the strong ceteris paribus conditions that exist across both years, how students answered these questions provides evidence as to whether their performance was affected by how the lectures were delivered: in-person or online.

It is to that difference we now turn, discussing first our data sources.

3 | THE HARD EVIDENCE: DATA FROM TEST SCORES

The primary independent variable in this study is student performance on 61 common final exam questions. This independent variable is paired with math SAT scores and several demographic variables for each student. Of these variables, the most important variable is whether the course was taken in 2019 (in-person) or 2020 (online). Other demographic variables include class year, college, or school of students' enrollment within the University of Virginia, gender, ethnicity, in-state residency status, international student status, and first-generation college student status.[8]

Final exam performance data were obtained from Elzinga's records as the instructor of the course. In 2019, the final exam was administered using hard-copy exam books and bubble sheets. These bubble sheets were not returned to students. As a result, we were able to rescore these bubble sheets to determine how many of the 61 common questions each student answered correctly. In the Fall of 2020, the final exam was administered using McGraw-Hill's *Connect* software with Proctorio proctoring extensions enabled.[9] These proctoring extensions create an environment consistent with the environments Bilen and Matros (2021) find limit cheating on online exams. *Connect* offers numerous analytic reports on student performance, which allowed us to determine how many of the 61 common questions each student answered correctly.

Math SAT scores were obtained from the University of Virginia's Student Information System (SIS). The school accepts both SAT and ACT scores for admissions, but the majority (74.4%) of students submit SAT scores.[10] When only an ACT score was submitted, we used the ACT/SAT concordance table published by the ACT to convert scores into the SAT scale. In cases where multiple test scores were available, we used the most recent SAT score. The left panels of Figure 1 show the distribution of SAT scores in 2019 and 2020.

We include math SAT scores as a control for unobservable student ability.[11] Previous studies have found that math standardized test scores, such as the ACT and SAT, are significant in predicting student performance in principals of microeconomics (Ballard & Johnson, 2004; Elzinga & Melaugh, 2009). Additionally, such standardized test scores are important controls in

[8]All data were accessed and used in compliance with UVA IRB-SBS protocol #4193. Under this protocol we attained passive consent from students to use their data in this study. Only three students elected to opt out of the study and are not included in this data set or analysis.

[9]The version of Proctorio used was the mid-level access version, which came included with the online textbook (and all McGraw-Hill eBooks copyrighted in 2019 or later). For additional details see Data S1.

[10]There were 37 cases where no standardized test score was available for a student. We omit these students from our sample.

[11]We do not use GPA as a proxy for ability for two reasons. First, we only have data on current GPAs, which include the course grade. Second, the majority of the students in the class are in their first semester of college, and thus do not have a GPA.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

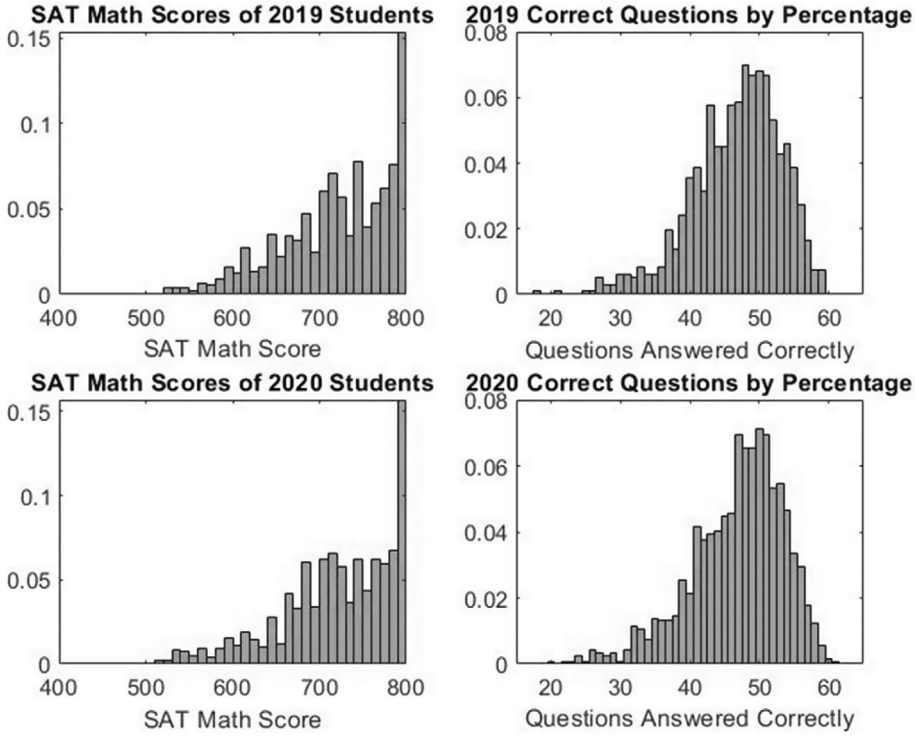

FIGURE 1 Distributions of SAT math scores and number of correct questions on course final exam. Left panels are distributions of SAT math scores. Right panels are distributions of correct questions. Top panels are distributions in 2019. Bottom panels are distributions in 2020. Distributions are shown normalized by percentage.

understanding the effects of other variables on student performance, which is the aim of this paper (Brown & Carl, 2002; Durden & Ellis, 1995).

A summary of the distribution of students into the demographic data and the average and standard deviation of correctly answered common exam questions for each demographic are shown in Table 1.[12] The demographics data was obtained from SIS under authorization of university administrators. Data on school, class year, and residency are from the date of the report, which was taken at the end of the Fall 2020 semester. However, the defined categories for school and residency rarely change once a student enters the University and it is simple to convert class year for students from 2019. Data on ethnicity and first-generation college student status are self-reported by students. At the 95% confidence level, we find no significant difference in the proportion of students in any demographic. At the 99% confidence level, we reject the null hypotheses that Engineering and Non-White/Hispanic students perform the same across both years. We find that on average both groups answered more questions correctly in 2020, under online instruction, than in 2019, under in-person instruction. In Section 3.2, we investigate these differences more thoroughly in a regression controlling for math SAT score and estimating the effects of all demographics jointly. These differences are no longer significant in this more rigorous estimation.

[12]In addition to the data displayed in Table 1, withdrawals after the withdraw date are similar. In 2019 there were 10 withdrawals, and in 2020 there were 10 withdrawals.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Summary of demographics.

		Proportions			Questions correct		
Category	**Demographic**	**2019**	**2020**	**Diff.**	**2019**	**2020**	**Diff.**
Number of students	–	960	1221	2181[a]	–	–	–
Gender	Female	0.5313 (0.0161)	0.5217 (0.0143)	−0.0095 (0.0215)	46.3765 (0.2827)	45.9451 (0.2679)	−0.4314 (0.3894)
School	Engineering	0.0896 (0.0092)	0.0934 (0.0083)	0.0038 (0.0124)	48.2791 (0.5387)	50.3246 (0.5159)	2.0455** (0.7459)
Class year[b]	First year	0.6169 (0.0158)	0.6495 (0.0137)	0.0325 (0.0209)	47.5334 (0.2514)	47.3153 (0.2282)	−0.2182 (0.3396)
Ethnicity	Non-White/Hispanic	0.3729 (0.0156)	0.3980 (0.0140)	0.0251 (0.0210)	45.3827 (0.3735)	46.7407 (0.3178)	1.3581** (0.4904)
First-generation	–	0.0802 (0.0088)	0.0901 (0.0082)	0.0099 (0.0120)	42.1948 (0.9210)	44.3727 (0.6352)	2.1779 (1.1188)
Residency	VA resident[c]	0.5854 (0.0159)	0.6265 (0.0138)	0.0411 (0.0211)	46.2206 (0.2700)	46.8301 (0.2361)	0.6094 (0.3586)
	International student[d]	0.0469 (4.7%)	0.0426 (0.0058)	−0.0043 (0.0089)	47.1111 (0.8612)	47.2308 (1.0051)	0.1197 (1.3236)
Test scores	SAT score	716.7292 (2.0848)	715.9132 (1.8875)	−0.8160 (2.8123)	–	–	–
	Questions correct	46.7531 (0.2083)	46.9083 (0.1910)	0.1551 (0.2826)	–	–	–

Note: The third through fifth columns show the demographics of 2019 and 2020 classes listed by proportion/mean. Standard errors are reported in parentheses bellow each mean. Differences are the 2020 value minus the 2019 value. No differences are significant at the 95% confidence level. The sixth through eighth columns show the mean number of common final exam questions correct for each demographic group with standard errors reported in parentheses. There are significant average performance improvements for Engineering and Non-White/ Hispanic students under online instruction at the 99% confidence level. Note: Categories with only one demographic listed are binary categories with the classification of interest listed.

[a]This number is total students in sample.

[b]There are 15 less observations for the class year category than other categories. This is because 15 students from 2019 exited the University in a manner other than graduation, resulting in a data record without an academic year.

[c]We define Virginia residency as students qualifying for in-state tuition.

[d]We define international students as students with a home address outside the United States.

***$p < .001$;

**$p < .01$; *$p < .05$.

22258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The right panels of Figure 1 show the distributions of the number of correct common final exam questions in each year. To make these distributions comparable, they are reported in percentage form because there were 960 observations in 2019 and 1221 in 2020. The histograms in Figure 1 show some differences in the sample distributions. However, the medians are the same (48 correct questions), and the means differ by only 0.1552 questions (46.7531 and 46.9083 questions, respectively).

3.1 | Results

We use a Mann–Whitney test for an initial comparison of student performance in online versus in-person classes.[13] This tests the null hypothesis that both sample distributions were drawn from the same population distribution (that is, the method of instruction has no effect on the distribution of correct common final exam questions). We are unable to reject this null hypothesis at any traditional confidence level (p-value = .3617). In short, there is no evidence that the method of instruction affects learning outcomes.

The left panels of Figure 1 show that approximately 15% of our students scored a perfect 800 on the math SAT. This is a significantly higher proportion of students than at most universities. To demonstrate that this is not driving our results, we exclude these students from the sample, and consider the Mann–Whitney test for the new distributions of correct final exam questions, shown in left panels of Figure 2. Our results from this test are unchanged from the results for the full sample, as we are still unable to reject the null hypothesis (p-value = 0.5858). We additionally test for a difference across instructional method in the distributions of correct questions for students with an 800 math SAT score (distributions shown in the right panels of Figure 2) and fail to reject the null hypothesis (p-value = 0.2937).

This non-parametric approach compares student performance only across the method of instruction and does not control for any other factors that may influence performance, including math ability. To control for math ability, we consider a baseline regression which regresses the number of common final exam questions a student answered correctly on the student's method of instruction and math SAT score. The results of this regression are reported in the second and fourth columns (labeled Baseline) of Table 2. The regression specification below shows the regression estimated in the SAT-online interaction columns of Table 2.

$$Questions\,Correct_i = \beta_0 + \beta_1 Online_i + \beta_2 MathSAT_i + \beta_3 MathSAT_i * Online_i + \epsilon_i.$$

The large number of students scoring an 800 on the math SAT indicates that this maximum score is truncating the distribution of math abilities for which the score is serving as a proxy. To ensure that this truncation is not affecting our results we run our baseline regression on the full data set and exclude students with a perfect math SAT score.

Using either data set, this baseline regression also finds no evidence that method of instruction affects student performance. For the full data set the coefficient for taking the class online indicates that, on average, students who took the class remotely answered 0.1975 more common questions correctly, after controlling for math SAT score. This is an insignificant difference. For

[13]For an explanation of the Mann–Whitney test and applications to small samples in experimental economics see Davis and Holt (1993). For our application we use the z-score approximation for large sample sizes, as derived in Mann and Whitney (1947).

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

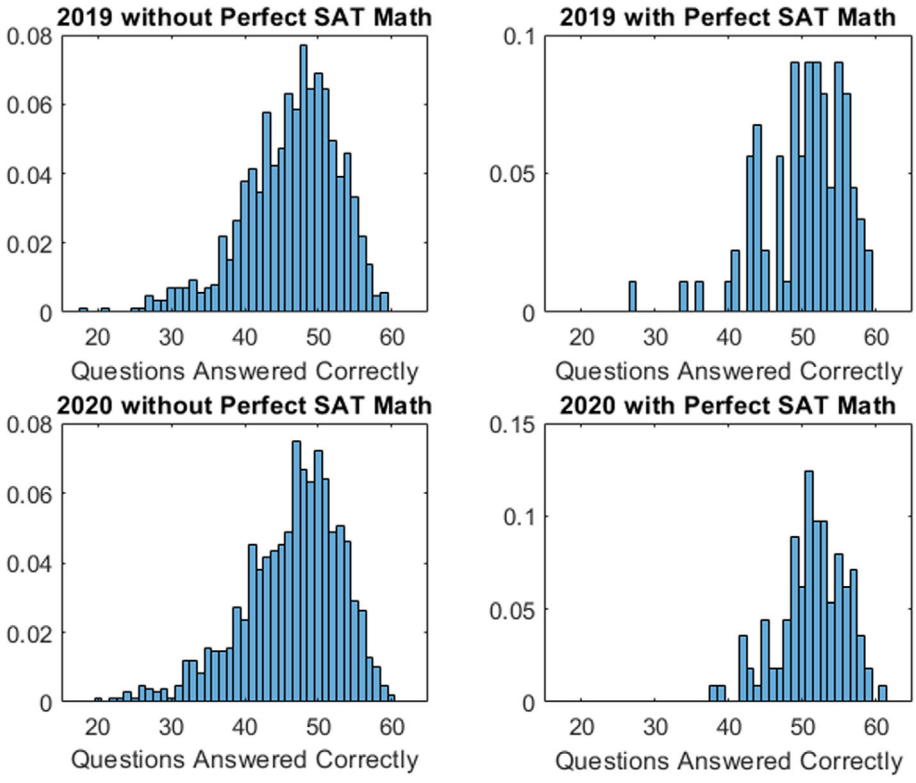

FIGURE 2 Distributions of correctly answered final exam questions, divided by math SAT score. The left panels show the distributions for students who did not score a perfect score. The right panels show the distributions for students with a perfect math SAT score of 800. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Baseline regressions results.

Data included	**All students**		**Students with SAT < 800**	
Specification	**Baseline**	**SAT-online interaction**	**Baseline**	**SAT-online interaction**
Intercept	9.5542*** (1.3367)	11.8681*** (2.0274)	9.2124*** (1.4595)	11.0253*** (2.2180)
Online	0.1975 (0.2432)	−3.8591 (2.6839)	0.1007 (0.2575)	−3.0707 (2.9330)
Math SAT score	0.0519*** (0.0018)	0.0487*** (0.0028)	0.0525*** (0.0020)	0.0499*** (0.0031)
SAT*online		0.0057 (0.0037)		0.0045 (0.0041)
R^2	0.2660	0.2668	0.2505	0.2510
N	**2181**		**1959**	

Note: Baseline Regression Results. Standard errors are reported in parentheses underestimates. Baseline regressions first control for math SAT score alone, and then also interact math SAT score with instructional method. This is done for the full data set (columns 2 and 3) and only for students with a math SAT score less than 800 (columns 4 and 5).
***$p < .001$; **$p < .01$; *$p < .05$.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

this reason, we cannot reject the null hypothesis that the coefficient is equal to zero. As a final robustness check, we estimate the impact of online learning on students at each quartile of the SAT distribution. The results of this check are included in Appendix A and we find no significant differences in any quartile.

Math SAT score has a significant impact on the number of questions a student answered correctly. For this reason, at the 99.9% confidence level we reject the null hypothesis that math SAT scores have no effect on performance. While a coefficient of 0.0519 may seem small at first glance, a couple examples suggest otherwise. The full data set baseline regression predicts that a student who scored a 500 on the math SAT (the national average) would answer approximately 36 questions correctly, whereas a student scoring an 800 would answer approximately 51 questions correctly: a difference of 15 questions or several letter grades on the final exam grading scale.[14]

Online education and testing create obvious opportunities for cheating that are not present in a proctored in-person exam. As discussed in the previous section, we attempted to limit cheating through the use of Proctrio's proctoring software. If cheating were widespread under online instruction, but not in-person instruction, mathematical ability would be a weaker predictor of performance in the online course. To test for this, we interact math SAT score and our indicator for online instruction in an augmented baseline regression. The results of this regression are reported in the third and fifth columns of Table 2. Both for the full and restricted data set, the interaction effect is positive and not significantly different from zero at the 95% confidence level. This is evidence against significant cheating under online instruction.

3.2 | Demographic groups

Students in different demographic groups might be affected differently by the two instructional methods being considered here. If some groups benefit from online learning, while other groups learn less, the lack of aggregate differences could be the result of these differences washing out across the sample. To account for these differences, we augment the baseline regression by including indicator variables for the demographic characteristics described in Table 1.[15] The results of this regression are reported in the fourth and fifth columns of Table 3. The regression specification for the fourth and fifth columns of Table 3 is included below. The specification for the second and third columns is the same, except it excludes the interaction terms.

$$\begin{aligned} \textit{Questions Correct}_i = {} & \beta_0 + \beta_1 \textit{Online}_i + \beta_2 \textit{MathSAT}_i + \beta_3 \textit{Female}_i + \beta_4 \textit{Female}_i * \textit{Online}_i \\ & + \beta_5 \textit{Engineering}_i + \beta_6 \textit{Engineering}_i * \textit{Online}_i + \beta_7 \textit{Non-White}_i + \beta_8 \textit{Non-White}_i \\ & * \textit{Online}_i + \beta_9 \textit{First-Generation}_i + \beta_{10} \textit{First-Generation}_i * \textit{Online}_i \\ & + \beta_{11} \textit{Out-of-State}_i + \beta_{12} \textit{International}_i + \beta_{13} \textit{Out-of-State}_i * \textit{Online}_i \\ & + \beta_{14} \textit{International}_i * \textit{Online}_i + \beta_{15} \textit{First-Year}_i + \beta_{16} \textit{First-Year}_i * \textit{Online}_i + \epsilon_i. \end{aligned}$$

[14]The grading scale for both the final exam and the overall class varies across years in order to maintain a stable distribution of letter grades. For the final exam scale, the only differences between the 2019 and 2020 scales came in minor adjustments necessitated by converting between a 180 point exam and a 200 point exam. For the overall course grading scale, there were some small differences mainly in increasing the thresholds for grades in the A and B range. These changes were necessitated by slightly higher midterm exam scores, which we believe were higher because of the inability to ask graphing questions in the online format. The grading scales from both years are available in Appendix B.

[15]This regression has 15 less observations due to the incomplete records of the academic year described above.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Demographics regression results.

	Demographics		Demographics with interactions	
Variable	**Coefficient**	**Standard error**	**Coefficient**	**Standard error**
Intercept	11.3906***	1.4436	11.7601***	1.4727
Online	0.1602	0.2426	−0.3470	0.5823
SAT score	0.0499***	0.0020	0.0498***	0.0020
Female	−0.1770	0.2473	0.0190	0.3681
Female*Online			−0.3735	0.4923
Engineering	0.4865	0.4468	−0.3499	0.6575
Engineering*Online			1.4824	0.8687
Non-White/Hispanic	−0.9610***	0.2676	−1.5403***	0.4046
Non-White/Hispanic*Online			1.0293	0.5401
First-generation	−1.0684*	0.4468	−1.5292	0.6858
First-Generation*Online			0.6946	0.8900
Out-of-state resident	−0.4526	0.2626	0.1642	0.3888
International	0.0574	0.6344	−0.0289	0.9410
Out-of-state*Online			−1.0790*	0.5228
International*Online			0.1823	1.2724
First-year	0.4556	0.2594	0.0294	0.3872
First-year*Online			0.8421	0.5144
R^2	0.2762		0.2821	
N	**2166**			

Note: Demographic Regression Results. Columns 2 and 3 show results when controlling for demographics, but not allowing demographics to have differential effects across instructional method. Columns 4 and 5 show results for demographics when allowing for demographics to have differential effects across instructional method.
***$p < .001$; **$p < .01$; *$p < .05$.

In this regression, the intercept, combined with the math SAT coefficient, predicts the score for a white male American in-state second-year (or above) student who is not in the School of Engineering nor is a first-generation college student. Such a student is predicted to answer 0.3470 less questions correctly when taking the class online versus taking the class in-person. However, this predicted difference cannot be differentiated from zero at a 95% or above confidence level. To ensure that the interaction terms between demographics and instructional method are not suppressing the statistical significance of this result, we include a regression specification with demographic controls but no interaction terms. The results of this regression are shown in the second and third columns of Table 3. We again fail to reject the null hypothesis that there is any difference in performance between online or in-person instruction.

By interacting the demographic variables with the method of instruction, we quantify the effects of online versus in-person instruction on individual demographic groups. For each demographic group, the difference in performance between online and in-person instruction is the combination of the online coefficient and the interaction coefficient, online + demographic*online. To test if this difference is significantly different from zero, we

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Impact of online versus in-person performance by demographic.

Demographic	Online differential	Test statistic
Female	−0.3280	1.7541
Engineering	1.1354	1.6115
Non-White/Hispanic	0.6823	1.0832
First-generation	0.3476	0.1179
Out-of-state	−1.4260*	5.1557
International	−0.1647	0.0129
First-year	0.4651	0.9474

Note: Online differential is the number of additional questions a student in a demographic group is predicted to answer correctly if they were to take the class online rather than in-person.
***$p < .001$; **$p < .01$;
*$p < .05$.

perform a joint hypothesis test for the addition of these coefficients. This effect for each demographic group and the corresponding chi-squared test statistic is shown in Table 4.

We find that out-of-state students perform differently between the two instructional methods. Holding all other factors constant, out-of-state students answer 1.4260 less questions correctly when taking the class online versus in-person. This is a relatively small difference out of 61 common questions but is sufficient to affect the letter grade of a student at the margin between two grades. At the 95% confidence level, we reject the null hypothesis that out-of-state students perform the same in-person and online.

One possible explanation for this difference is unobserved variable bias. In the Fall 2020 semester students were allowed to live on campus but there were no residency requirements (as is normally the case for first-year students). However, all students were required to return home for final exams and winter break. We were not permitted to access records on which students returned to campus (or Charlottesville) during the semester. There may be a negative effect to taking online classes from a student's home rather than being on campus. If out-of-state students were less likely to be in residence during the COVID-19 pandemic, their performance loss may be due to this effect rather than their out-of-state status.

Overall, there is limited evidence that different demographic groups perform differently depending on the method of instruction. While out-of-state students experience a negative performance difference with online instruction, the difference is small, and amounts to approximately 1% of a student's final cumulative grade. The general finding that online learning does not affect student performance is consistent across most demographic groups.

3.3 | Questions divided by skill type

In addition to affecting a student's overall learning in a class, online learning may affect how a student learns to solve different types of problems. For example, the ability to rewatch recorded lectures in an online class may enhance a student's comprehension of the vocabulary of economics but the lack of personal access to the professor may impair learning how to apply economics in a problem-solving manner. To understand how online and in-person learning affect a student's ability to solve different types of economic problems, we divide the common final

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

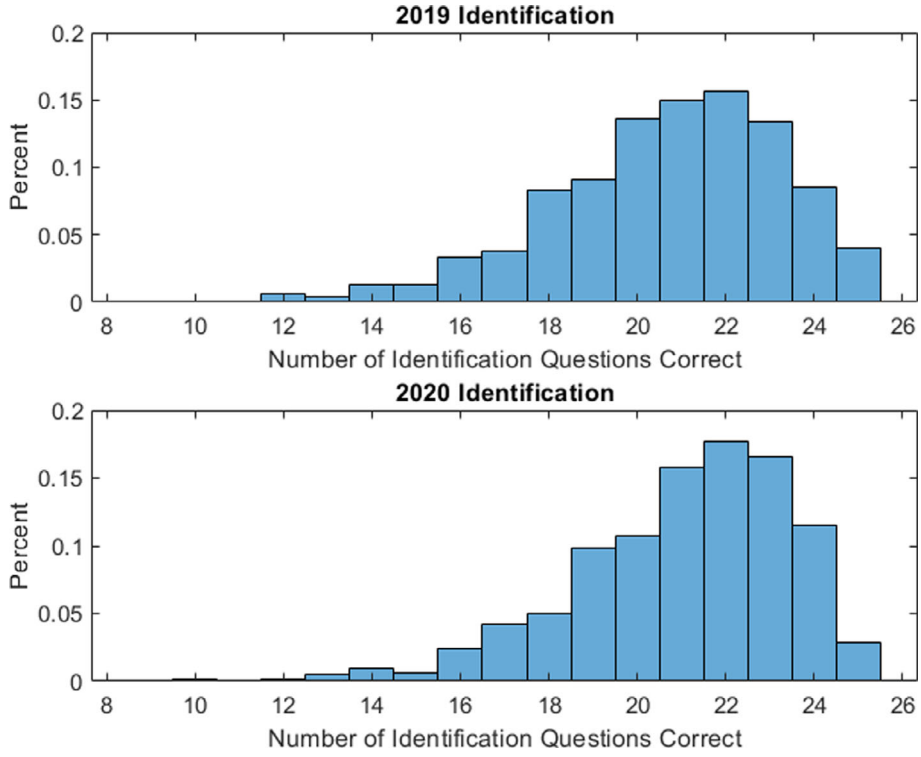

FIGURE 3 Histograms of distribution of correct identification questions by year. The top panel shows the distribution for in-person instruction in 2019. The bottom panel shows the distribution for online instruction in 2020. Distributions are normalized to percentage terms. [Color figure can be viewed at wileyonlinelibrary.com]

exam questions into three categories: identification, non-quantitative analysis, and quantitative analysis.

Identification questions require a student to identify the meaning of an economic term or use a very simple application of the definition that is not substantively different than recalling the definition. Of the 61 common questions, 25 were identification questions. Quantitative analysis questions require a student to recall and manipulate a formula or apply some calculation to answer the question. Six of the 61 common questions were of this sort. Non-quantitative analysis questions are designed to test economic intuition. These questions require students to apply economic reasoning to solve a problem that does not require the use of arithmetic. This type of question occupies the largest share of the common questions, 30 out of 61.

Figures 3–5 show the distribution of the number of questions of each type students answered correctly when receiving in-person (2019) and online (2020) instruction.[16] We begin our analysis of the question-type results at a non-parametric level, using Mann–Whitney tests. In Figure 3, the distribution of correct identification questions under online instruction appears to have a greater mass in the right of the distribution than the distribution under in-person instruction. The Mann–Whitney test rejects the null hypothesis that these distributions are drawn from the same population distribution at the 99.9% confidence level ($p = .0005$). This

[16]Across both years, 18 students took their exam under unique circumstances, which now limit our ability to obtain question level results for these students. They are excluded from this cognitive type analysis.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

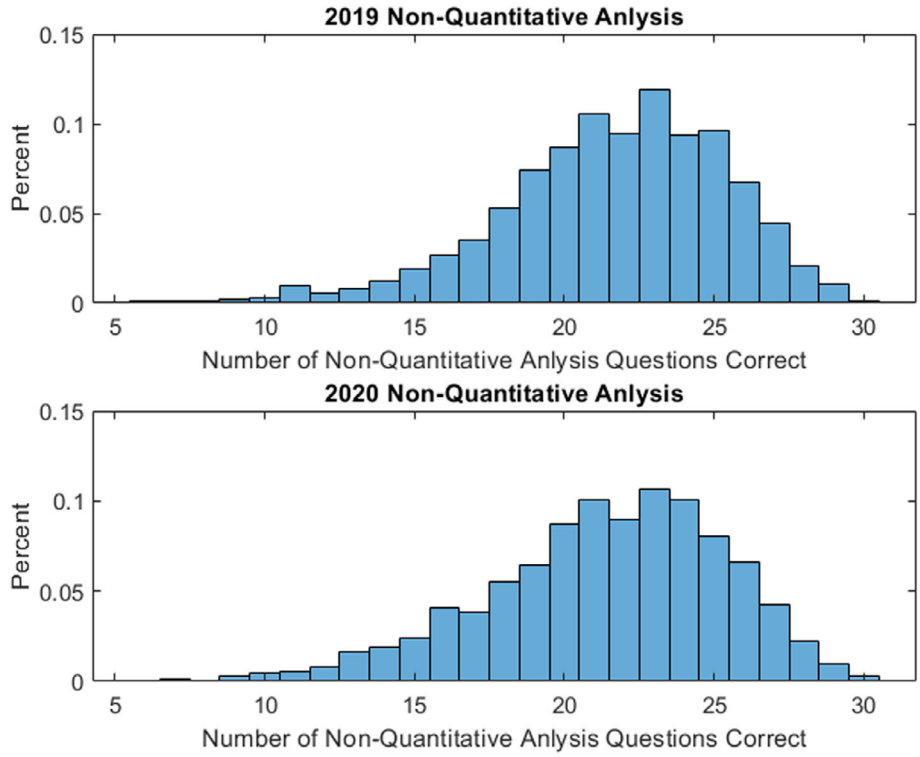

FIGURE 4 Histograms of distribution of correct non-quantitative analysis questions by year. The top panel shows the distribution for in-person instruction in 2019. The bottom panel shows the distribution for online instruction in 2020. Distributions are normalized to percentage terms. [Color figure can be viewed at wileyonlinelibrary.com]

provides some evidence that online instruction improves student's performance on identification questions.

Figure 4 shows the distributions of correct non-quantitative analysis questions. These distributions do not appear substantially different, and we cannot reject the null hypothesis at the 90% confidence level ($p = .1237$). The distributions of correct quantitative analysis questions are shown in Figure 5. These distributions show a higher percentage of students receiving online instruction answering 5 and 6 questions correctly. At the 95% confidence level ($p = .0213$) we are able to reject the null hypothesis that these two distributions are drawn from the same population distribution.

While these non-parametric results provide insight into differences in student performance, we use regressions similar to our earlier baseline regression analysis for a more robust analysis. Table 5 shows the results of regressing the number of questions answered correctly for a given type on method of instruction and math SAT score. This is not accomplished in the non-parametric tests discussed above. The regression specification used for each type of question is included below.

$$Questions\ Correct_i = \beta_0 + \beta_1 Online_i + \beta_2 Math\,SAT_i + \epsilon_i.$$

The results of these regressions are broadly consistent with the results of the non-parametric tests. That is, online learning positively impacts performance on identification and quantitative

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

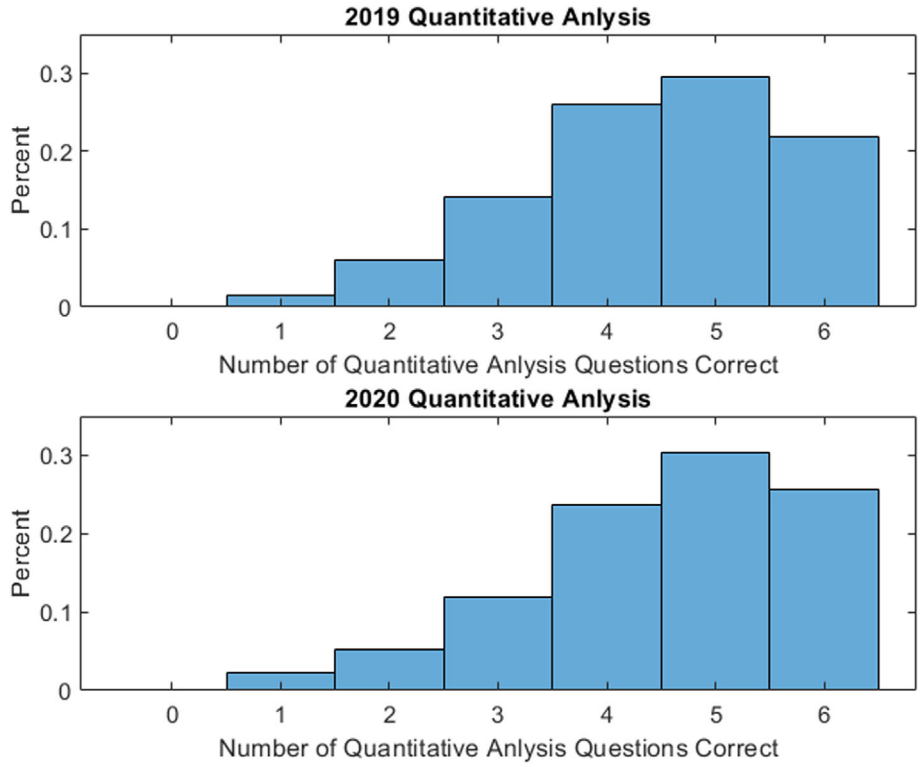

FIGURE 5 Histograms of distribution of correct quantitative analysis questions by year. The top panel shows the distribution for in-person instruction in 2019. The bottom panel shows the distribution for online instruction in 2020. Distributions are normalized to percentage terms. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Regressions using questions divided by cognitive type.

	Identification		Non-quantitative		Quantitative	
Independent variable	**Coefficient**	**Standard error**	**Coefficient**	**Standard error**	**Coefficient**	**Standard error**
Intercept	11.064***	0.5724	0.7146	0.8128	−2.2481***	0.2632
Online	0.3815***	0.1039	−0.2744†	0.1476	0.1083*	0.0478
Math SAT	0.0134***	0.0008	0.0292***	0.0011	0.0093***	0.0004
R^2	.1219		.2397		.2338	

***$p < .001$; **$p < .01$; *$p < .05$; †$p < .1$.

analysis questions. We estimate that students taking the class online answer 0.3815 *more* identification questions correctly than those taking the class in-person. At the 95% confidence level, we reject the null hypothesis that this effect is zero. Similarly, we estimate a difference of 0.1083 *more* correct quantitative analysis questions for students taking the class online. We again reject the null hypothesis that this difference is zero at the 95% confidence level. These are relatively

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

small improvements from online learning—both are less than one additional correct question—but do show statistically significant gains from online learning.

We find a point estimate of 0.2744 *less* correctly answered non-quantitative analysis questions when receiving online instruction. This difference is statistically significant at the 90% confidence level, where we can reject the null hypothesis that there is no difference between instructional methods. This provides weaker evidence that online learning leads to relatively small losses in learning how to apply economic intuition.

One possible explanation for these results is that online learning allows students to rewatch lectures and creates additional barriers to students interacting with the professor. Identifying definitions and manipulating formulas are skills that are likely improved by rewatching a lecture where the terms or formulas are introduced. These skills directly translate to answering identification and quantitative analysis questions. After an in-person lecture (or discussion section) students can approach the instructor (or TA) to ask clarifying questions or questions that attempt to connect different concepts, without having to draft an email or attend office hours. This ability to ask questions easily may help students improve their economic intuition and score better on non-quantitative analysis questions.

3.4 | Grading basis

One factor that may bias these results is whether students took the course on a Credit (CR)/General Credit (GC)/No Credit (NC) grading basis, rather than for a letter grade. In 2019 all students in the sample took the course for a letter grade. Because of COVID-19, Fall 2020 students could choose a CR/GC/NC grading basis for any of their courses between November 2 and November 6, 2020, 3 weeks before the end of the semester. This new policy also permitted grades of CR in Fall 2020 courses to count toward degree requirements. In Econ 2010, 222 students (or 18.18%) elected to take the course for CR/GC/NC in Fall 2020.

A student's choice to take the course for credit and not a grade was likely endogenous to that student's performance in the course. Students already had completed both midterms by the decision date and thus could make a reasonable prediction as to their final exam performance. Students who thought they were performing poorly in the course had an incentive to take the credit only option, which did not affect their GPA. If this were the only endogeneity associated with this grading option, it would not be a major concern. However, students taking the course only for credit may have been incentivized to spend less time studying for the final exam. In this case, they would likely perform worse on the final exam than if they were taking the course for a letter grade. This would bias our results by reducing scores in the online treatment.

To address this concern, we compare the distributions of students' scores on the common final exam questions between students who took the class for a grade versus for credit only in 2020. Using a Mann–Whitney test we are able to reject the null hypothesis that these students' scores are drawn from the same distribution at the 99.9% confidence level. We find that students taking the course on a non-graded basis performed worse on the exam than those taking the course for a grade. They correctly answered approximately 7 fewer questions on average. To distinguish whether this difference is due to lower performing students choosing to take the course for credit only, or students who chose to take the course only for credit being less motivated and thus performing worse, we use a Probit model to estimate the probability of a student taking the course for credit only.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The primary independent variables of interest in our analysis of the decision to take the course only for credit and not a grade is a student's performance on the midterm exams. Two such exams were given during the course. Both were short answer exams scored on an 80 point scale, with a bonus point available for solving a riddle (maximum score = 81). Midterms were graded by a student's TA using a common answer key. To maximize the consistency of grading across TAs, the keys were each reviewed and discussed with all TAs in 2+ hour training meetings.

Students received their grades for both midterm exams by the final day to select into the credit-only grading scheme.[17] If students with lower midterm grades were more likely to take the course for credit only, this would indicate that at least some of the difference in final exam scores is explained by the fact that lower-performing students were more likely to take the course for credit only. In columns 2 and 3 of Table 6 we report the results of a Probit regression prediction of the probability of a student taking the course only for credit, based on their midterm grades.

We find that both midterm grades are significant in predicting a student's decision to take the course on a non-graded, credit-only basis at the 99.9% confidence level.[18] Our estimates of both coefficients are negative, meaning that increasing a student's grade on either midterm decreases the probability of students taking the course for credit only.[19] Put differently, students who performed poorly on their midterms were more likely to take the course only for credit. To better illustrate these differences, we consider the predicted probabilities of students selecting to take the course only for credit at selected midterm grades. A student earning the class average on both midterms (a 66.68/80 on Midterm 1 and a 62.90 on Midterm 2) has a predicted probability of taking the course only for credit of 20.86%. For a student who earned a perfect score on both midterms (80/80), the predicted probability of that student taking the class only for credit falls to 2.01%. For a student struggling in the class (scoring a 55/80 on both midterms), the probability of taking the class only for credit rises to 37.96%. Thus, our hypothetical struggling student is 17.10% more likely to take the course only for credit than the average student, and 35.95% more likely than a top performing student. This supports the view that students taking the course only for credit were under-performing students who still needed to put effort into the final exam in order to pass.

While performance on midterm exams is a significant predictor of a student's choice of grading option, other demographic factors also may affect this decision. To investigate this concern, we include our demographic controls in the Probit regression reported in columns 4 and 5 of Table 6. The most interesting result from this demographic regression is that first year students

[17]A small number of students took a make-up exam to replace a midterm missed due to illness or a varsity athletic event. The make-up exam was given after the final day to opt into the credit only grading scheme. We are unable to eliminate these students from the data set, so these make-up exam scores are included as regular mid-term scores. Additionally, some students appealed their midterm grades and received points back due to grading errors, or acceptable (but unique) answers that were not on the original key. The number of students receiving these points was very small, but the appeals process for the second midterm was not closed until after the date to select into credit only grading. Our data set only includes the midterm grades, after appeals.

[18]Although our point estimates of the coefficients differ for midterms 1 and 2, these differences are not statistically significant.

[19]As a robustness check we divide the students into quartiles based on their combined score on midterms 1 and 2, and run the baseline regression. The full results of this robustness check are available in Appendix A. We find that students in the upper half of the distribution are much less likely to take the class for credit only, and the individual exam scores have no significant effect on this probability. For students in the lower half of the distribution, higher midterm scores have a significant effect on reducing the probability of a student taking the course for credit only.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Credit-only probit regression results.

	Baseline		Demographics	
Variable	**Coefficient**	**Standard error**	**Coefficient**	**Standard error**
Intercept	4.3700***	0.3493	4.7969***	0.3958
Midterm 1	−0.0398***	0.0048	−0.0427***	0.0050
Midterm 2	−0.0454***	0.0054	−0.0440***	0.0056
Female			−0.1152	0.1067
Engineering			0.2891	0.1662
Non-White/Hispanic			0.0837	0.1117
First-generation			−0.0626	0.1735
Out-of-state resident			0.0190	0.1145
International			0.2871	0.2582
First-year			−0.6220***	0.1064

Note: Columns 2 and 3 show results for baseline regression using grades for midterms 1 and 2. Columns 4 and 5 show results for demographics regression which augments the baseline regression with the described demographic variables.
***$p < .001$; **$p < .01$, *$p < .05$.

were significantly less likely to opt to take the course on a credit-only basis. At the 99.9% confidence level, we are able to reject the null hypothesis that first years are equally likely to take the course only for credit as upper-class students. For example, an "average" upper-class student is predicted to take the course only for credit with a probability of 25.91%.[20] The same "average" first-year student is predicted to have an 11.98% probability of taking the course only for credit; a 13.93% reduction.

One possible explanation for this difference is that first-year students planning to apply to the School of Commerce or the School of Leadership and Public Policy chose to take the course for a grade. As is discussed in Section II, admission to these schools is competitive. Both schools announced that they would accept students who chose to take the prerequisites only for credit during the COVID-19 pandemic. However, conversations with students revealed that they were unsure that a grade of CR would be treated the same as an A or B. This would imply that first-year students were less likely to take the course only for credit not because of their performance in the course, but because of external concerns. Another possible explanation for the difference is that upper-class students are more likely to take the course to fulfill a general education requirement. Some of these students may have wanted to put less effort into the course, and thus chose to take the course only for credit.

In examining students' demographic profiles and their performance up to the deadline for choosing to take the course on a graded or non-graded basis, we find that performance and demographics matter in a student's decision to take the course only for credit. Because factors other than performance are significant in the prediction of this decision, we cannot rule out the possibility that some students taking the course only for credit exerted less effort on the course,

[20]We define an "average" student in the context to be a student scoring the class average on both midterms (66.68/80 and 62.90/80) and being of the modal type in each demographic category. Thus, they are female, not in engineering, white, and in-state.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

skewing down our results for the online treatment. This means that our results are lower bounds on the effect of online instruction.

3.5 | Discussion

Our results indicate that in terms of student performance on exams, there is no overall difference between online and in-person instruction, though some differences exist for out-of-state students and across questions requiring different cognitive skills: on the positive side, there is evidence that online learning improves performance on identification and quantitative questions, and on the negative side, there is evidence that online learning reduces the performance of out-of-state students and weaker evidence it reduces performance on non-quantitative questions. These differing results make up some of the costs and benefits of online learning.

The differences in performance across question type pose the question of what skills instructors want students to take away from an economics class? If the goal is to teach students the vocabulary of economics or how to use key formulas from economics, then online learning provides a marginally better environment for learning these skills. On the other hand, if the goal is to teach students economic reasoning and intuition then our results indicate that online learning may be a marginally worse instructional method than in-person instruction. Thus, the value an instructor places on these different skills may sway whether the costs outweigh the benefits of online learning.

These costs and benefits to student performance must be balanced with the pecuniary and administrative costs and benefits of online learning experienced by colleges and universities and their faculty. In this case, the short-run marginal cost of both instructional formats was negligible. The University of Virginia already owned both a large lecture hall and a fully equipped film studio. However, the difference in long-run marginal cost of these instructional formats may be considerable and are beyond the scope of this paper. The administrative costs (mainly instructor time) were higher for online instruction than for in-person instruction, even when accounting for the differences in class size.[21] However, these administrative costs may diminish upon repetition of an online class. The weight to be placed on these costs and benefits also is beyond the scope of this paper.

An obvious question that flows from our research is: How general are these results? The results would apply to any conventional introductory course in economics where the topics covered involve the standard fare of production possibility curves, demand and supply, elasticity, theories of competition and monopoly, labor markets, and income distribution. Beyond that, they should be applicable across disciplines at the college and university level for any course typically taught in a "chalk and talk" format. That is, we know of no reason that our findings for teaching economics would not be relevant for introductory courses in the natural sciences (such as Chemistry); in business (such as Accounting); and in the social sciences (such as History). We are less optimistic that our results apply in the Humanities, where the large lecture format is less common.

We are not persuaded that our results have bearing upon the delivery of education at the elementary, middle school, and high school levels. The conventional wisdom is that live and in-person education is more important for young students, who do not have the maturity or the

[21]This difference could be the result of Elzinga's experience with in-person instruction and inexperience with online instruction.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

skills to concentrate for hours in front of a computer. This is reinforced by Halloran et al. (2021) who find that online learning during COVID-19 has negatively impacted test scores at this level of education.

While many of the conditions line up to allow a comparison between online versus in-person instruction, we recognize that disentangling the effects of the change in method of instruction from the effects of the pandemic itself is difficult. A methodological irony of our research is that the event that allows us to make our comparison in instructional methods—the COVID-19 pandemic—may itself taint the results we find because of consequences we cannot observe or measure.

In short, our research does not account for whether there may be alternative explanations—provoked by the pandemic itself—that affect our results. One explanation that comes readily to mind is whether students devoted additional study time during the pandemic because constrained extracurricular opportunities reduced the opportunity cost of time spent studying. Alternatively, because so many classes (in addition to Econ 2010) had to be taken remotely, students may have suffered "Zoom fatigue" and this reduced time devoted to study. Another variable that could taint conventional measures of student performance, distinct from actual learning, is whether there is a difference in the propensity of students to cheat as between online versus in-person instruction.

Because of the size of our data set, the variables held constant, and the stark, binary difference between in-person instruction and online instruction, we place most of our confidence in our empirical analysis of test results. However, there also is soft evidence that merits examination. This comes in the form of student evaluations of the same course taught in two different ways. It is to that evidence that we now turn.

4 | THE SOFT EVIDENCE: DATA FROM COURSE EVALUATIONS

College administrators routinely use course evaluations to evaluate the quality of a professor's teaching relative to other professors and to evaluate a professor's teaching effectiveness over time. Despite the widespread use of course evaluations, there is scant research proving that professors with high evaluations are effective teachers (Linask & Monk, 2018). Evaluations may be influenced by factors like expected grades (Boring et al., 2016), gender (Mitchell & Martin, 2018), ethnicity (Stark & Freishtat, 2014), whether the course is mandatory, the class size (Ragan Jr & Walia, 2010), and other variables.

In his survey of research on course evaluations, DeLoach (2012) concludes that a student's evaluation of teaching is different from a student's evaluation of what is learned in the course. Students generally can assess such variables as a professor's teaching skill, the professor's accessibility and rapport with students, and the level of course difficulty (Cashin, 1999) but are less capable of evaluating a professor's knowledge of the material and the quality of course design. That said, student evaluations are widely used (indeed, often mandated) to evaluate faculty over and above alternative methods such as peer review (Arreola, 2000; Chism, 1999). Because of their ubiquitous application to course assessment and data availability, we also make use of student evaluations to address the question of online versus in-person instruction.

The option to complete a course evaluation is open to all students taking undergraduate courses at the University of Virginia. The evaluation form is standard across courses, but professors can modify the form to fit the circumstances of the course. Students are invited (but not required) to complete the evaluation forms at the end of a course. Some professors offer a modest incentive for

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

completion. In Econ 2010, for example, students who completed the course evaluation for their discussion section received an extra credit point worth ¼% of their grade. This incentive did not directly carry over to the lecture course evaluation, but likely improved the response yield.

For this study, many of the variables that can affect course evaluations are held constant. By comparing evaluations of the same professor from different years, factors like relative popularity or gender are negligible. Similarly, because Elzinga has been teaching for decades, his relative effectiveness and performance should remain constant (though one would hope for improvements each year). To the extent that course evaluations provide an indirect measure of student learning, we analyzed this metric to further understand differences in effectiveness that might prevail between in-person versus online learning.

In reading student comments (rather than only comparing the numbers), one is left with the impression that students who took Econ 2010 remotely in 2020 wrote positive evaluations either because they had low expectations of an online course, or because they compared the online delivery of Econ 2010 with other online courses, which may not have been as engaging or polished. Examples of this sentiment can be seen in the following quotes pulled from actual student evaluations in 2020.

> "I thoroughly appreciated the effort put into making this class as close to in–person as possible, with the professionally recorded lectures and proctored exams."
> "I felt that the course was actually best adapted out of all my courses this semester. You're doing it right. Elzinga was very responsive to emails and the same with the TAs."

The second quote summarizes the idea that Econ 2010 was a well-adapted online course, and that students were directly, and favorably, comparing it to other online courses. Put differently, the soft evidence may not reveal that the online delivery of Econ 2010 was better than in-person delivery but simply that it was better than the average expectation for an online course.

Course evaluation data were obtained from the lecture sections of Econ 2010 for 2019 and 2020. Student evaluations of their TAs and individual discussion sections were not included. The 2 years of lecture sections offered a robust data set capturing student evaluations under conditions of in-person and online learning. The questions and interface of course evaluations between these two periods were not identical, but evaluations for both years were online and had similar question structures with Likert item, multiple choice, and open-ended questions.[22]

The 2019 and 2020 course evaluations share one identical question and three comparable questions. Three of these questions are relevant to this study: questions on availability, effectiveness, and amount learned. The questions regarding effectiveness and amount learned shared answer choices between the 2 years, while the question regarding availability had different options in 2019 than in 2020. Students were not required to answer all questions; the number of responses to each question aggregated across years and sections ranges from 1602 to 1608.

A number was assigned to each category: Strongly Agree received a 5, Agree received a 4, Neutral received a 3, and so on. Because the instructional method is nominal and the evaluation responses are ordinal, no normality assumptions can be made about the distribution of

[22]In both years the evaluations were hosted on an online platform students could access using a link in regular reminder emails. However, the university switched online platforms between the two semesters. In 2020, the number of questions was reduced from 33 to 16 and two conditional questions were added. Conditional questions are of the structure: "Because you answered Strongly Agree or Agree to question 5..."

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 Predicting course evaluation response—ordinal regression results.

Dependent variable	Online coefficient	Standard error	Number of responses
Availability	−0.606***	0.0942	1602
Effectiveness	0.453***	0.0982	1607
Amount Learned	0.634***	0.0981	1606

Note: Regression results from ordered logit model predicting course evaluation ratings based on method of instruction.
***$p < .001$.

responses, and the intervals between responses may not be equal.[23] Mann–Whitney tests are used to analyze distributions and an ordered logistic regression is used to analyze the effect of the year on the latent variable.

A priori, one might predict that course evaluations for classes taught online would be lower. Consistent with this, some studies cite online learning as an important factor in lower course evaluations, especially compared to in-person, traditional courses (Farinella, 2007). A reduction in instructor availability could follow a move to online instruction. This alone could restrain student enthusiasm for a course.

Using Mann–Whitney tests, the null hypothesis that the 2019 and 2020 samples have the same distribution is rejected for availability, effectiveness, and amount learned.[24] Based on these data, online learning significantly altered the distributions of responses to the course evaluation questions. The direction of these changes can be determined from the ordinal logistic regression coefficients seen in Table 7. These coefficients indicate that at the 99.9% confidence level students rated instructor availability as lower, but rated instructor effectiveness and amount learned higher in 2020 during the period of online learning, compared to during conditions of in-person teaching.

4.1 | Ceteris paribus concerns

Changes in the content of course evaluations are a threat to the all-else-equal assumption present in these results. The 2019 evaluation contained more questions than the 2020 version; this may have caused student respondents to spend less time reading and answering each question. In addition, the scope of students' comparisons is not necessarily constant. The course evaluations may reflect student opinions about Econ 2010 as taught online compared with other courses students were taking online, instead of an evaluation of the course on its own merits. This effect may have been intensified in 2020 because professors had varying ability to convert courses to an online format. If Elzinga succeeded in online teaching in comparison to the other professors of his students, the students may have been more inclined to rate Elzinga higher (even if his online course was equally as good as or worse than his in-person course). For this reason, we put more weight on the test data than we do on course evaluation data.

[23]For instance, the difference between Neutral and Agree may be larger than the difference between Agree and Strongly Agree.
[24]The *p*-value for each of these Mann–Whitney tests is approximately 0.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | CONCLUSION

In the recent past, the center stage issue in teaching economics has been the merits of passive versus active classroom pedagogy: do students learn more economics through the traditional lecture ("chalk and talk") or through active classroom engagement (such as cooperative learning exercises and classroom experiments). The COVID-19 virus of 2020–2021 has pushed to center stage the issue of online versus in-person classroom pedagogy. The underlying concern with student learning remains the same but the pandemic changed the question—and the pandemic has enabled more empirical research on the topic.

In addition to the question of student learning, student preferences and well-being are a vital component of evaluating online and in-person learning. Online learning is inherently more isolating than in-person learning. This may carry psychological and economic costs for students, particularly if student networking opportunities are limited by online learning. We largely leave this topic to future research. However, through our analysis of course evaluations we find some soft evidence that student's do not dislike all forms of online learning.

One swallow does not make a Spring. In like fashion, one comparison of student performance between in-person and online pedagogical formats does not make a definitive case. The academic cliché that "more research is needed" applies here. Nonetheless, our research indicates that overall, there is no difference in student test results between the 2019 and 2020 classes. However, there is evidence that students' ability to answer some types of questions improved from online learning (identification and quantitative analysis questions). On the other hand, there is evidence that some students experienced performance losses from online learning (out-of-state students). Therefore, overall, there are costs and benefits associated with online learning, at least from the perspective of student performance.

ACKNOWLEDGMENTS

The authors are grateful to Annamarie Black for data support, to Elizabeth Magill and Michael Citro for IRB compliance support, and to Craig Epstein, Jackson Howell, and Taylor Thiessen for research assistance. We thank three anonymous referees for their helpful comments. We are grateful to Anthony Underwood for his insightful discussion of our paper at the 2021 SEA Annual Meeting and Avi J. Cohen for his insightful discussion of our paper at CTREE 2022. We thank the Bankard Fund for financial support. We also thank Lee Coppock, Anton Korinek, John Pepper, Sarah Turner, William Wood, and members of the University of Virginia Quantitative Collaborative for helpful comments and Sarah Lapp for editorial assistance.

REFERENCES

Alpert, W., Couch, K.A. & Harmon, O. (2016) A randomized assessment of online learning. *American Economic Review: Papers and Proceedings*, 106(5), 378–382.

Arreola, R. (2000) *Developing a comprehensive faculty evaluation system: a handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system*. Bolton, US: Anker.

Ballard, C.L. & Johnson, M.F. (2004) Basic math skills and performance in an introductory economics class. *Journal of Economic Education*, 35(1), 3–23.

Bilen, E. & Matros, A. (2021) Online cheating amid COVID-19. *Journal of Economic Behavior and Organization*, 182, 196–211.

Boring, A., Ottoboni, K. & Stark, P.B. (2016) *Student evaluations of teaching (mostly) do not measure teaching effectiveness*. ScienceOpen Research.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Brown, B.W. & Carl, E.L. (2002) Can web courses replace the classroom in principles of microeconomics? *American Economic Review: Papers and Proceedings*, 92(2), 444–448.

Calafiore, P. & Damianov, D.S. (2011) The effect of time spent online on student achievement in online economics and finance courses. *Journal of Economic Education*, 42(3), 209–223.

Cashin, W.E. (1999) Student ratings of teaching: uses and misuses. In: Seldin, P. (Ed.) *Current practices in evaluating teaching: a practical guide to improved faculty performance and promotion/tenure decisions*. Bolton, US: Anker, pp. 24–44.

Chism, N. (1999) *Peer review of teaching: a sourcebook*. Bolton, US: Anker.

Davis, D.D. & Holt, C.A. (1993) *Experimental Economics*. Princeton, New Jersey: Princeton University Press.

DeLoach, S.B. (2012) What every economist should know about the evaluation of teaching: a review of the literature. In: Hoyt, G.M. & McGoldrick, K.M. (Eds.) *International handbook on teaching and learning economics*. Northampton, Massachusetts: Edward Elgar.

Durden, G.C. & Ellis, L.V. (1995) The effects of attendance on student learning in principles of economics. *American Economic Review: Papers and Proceedings*, 85(2), 343–346.

Elzinga, K.G. & Melaugh, D.O. (2009) 35,000 principles of economics students: some lessons learned. *Southern Economic Journal*, 76, 32–46.

Farinella, J. (2007) Professor and student performance in online versus traditional introductory finance courses. *Journal of Economics and Finance Education*, 6(1), 40–47.

Figlio, D., Rush, M. & Yin, L. (2013) Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, 31(4), 763–784.

Gratton-Lavoie, C. & Stanley, D. (2009) Teaching and learning principles of microeconomics online: an empirical assessment. *Journal of Economic Education*, 40(1), 3–26.

Halloran, C., Jack, R., Okun, J.C. & Oster, E. (2021) Pandemic schooling mode and student test scores: Evidence from US states. *NBER Working Paper* No. 29497.

Kofoed, M., Gebhard, L., Gilmore, D. & Moschitto, R. (2021) Zooming to class?: experimental evidence on college students' online learning during COVID-19. *IZA Discussion Paper* No. 14356.

Linask, M. & Monk, J. (2018) Measuring faculty teaching effectiveness using conditional fixed effects. *Journal of Economic Education*, 49(4), 324–339.

Mann, H.B. & Whitney, D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60.

Miller, J.D. & Rebelein, R.P. (2012) Research on the effectiveness of non-traditional pedagogies. In: Hoyt, G.M. & McGoldrick, K.M. (Eds.) *International handbook on teaching and learning economics*. Northampton, Massachusetts: Edward Elgar.

Mitchell, K.M.W. & Martin, J. (2018) Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648–652.

Orlov, G., McKee, D., Berry, J., Boyle, A., DiCiccio, T., Ransom, T. et al. (2020) Learning during the COVID-19 pandemic: it is not who you teach, but how you teach. NBER working paper No. 28022.

Ragan, J.F., Jr. & Walia, B. (2010) Differences in student evaluations of principles and other economics courses and the allocation of faculty across courses. *Journal of Economic Education*, 41(4), 335–352.

Sosin, K., Blacha, B.J., Agarwal, R., Bartlett, R.L. & Daniel, J.I. (2004) Efficiency in the use of technology in economic education: some preliminary results. *American Economic Review: Papers and Proceedings*, 94(2), 253–258.

Stark, P.B. & Freishtat, R. (2014) *An evaluation of course evaluations*. ScienceOpen Research.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Elzinga, K. G., & Harper, D. Q. (2023). In-person versus online instruction: Evidence from principles of economics. *Southern Economic Journal*, 1–28. https://doi.org/10.1002/soej.12635

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX A: QUARTILE REGRESSION ROBUSTNESS CHECKS

We divide students by rough SAT quartiles to verify that online instruction does not have a different effect on at risk students. The quartiles used in Table A1 are not precise divisions of the data into quarters. This is done to ensure that all students with the same Math SAT score are included in the same grouping for analysis. The number of students included in each regression is reported in the *N* rows. We find that online instruction does not have a significant impact on student performance in either the baseline or SAT interaction regressions for any quartile. The point estimates vary widely, but no estimates are significant.

To further explore the effects of midterm scores on a student's likelihood of taking the course for credit-only, we divide students into quartiles based on their combined score on the midterms. In dividing students we prioritized keeping all students with the same midterm score in the same estimation. For this reason, students are not evenly divided into quartiles, but the selected ranges are as close as possible to an even division without splitting across scores. We find that the most at risk students are the most likely to take the course for credit only. For these students, their second midterm has a significant effect on the likelihood of taking the course for credit-only but the first midterm does not have a significant effect. This could be the

TABLE A1 Baseline regression results by SAT quartile.

Data included	**SAT ≤ 670**		**670 < SAT ≤ 720**	
Specification	**Baseline**	**SAT-online interaction**	**Baseline**	**SAT-online interaction**
Intercept	−2.9424 (4.3818)	5.8888 (7.0557)	16.4319 (11.5791)	−3.2720 (18.2110)
Online	0.2405 (0.5816)	−14.0197 (8.9571)	−0.4991 (0.4699)	32.4981 (23.5563)
Math SAT score	0.0717*** (0.0070)	0.0576*** (0.0113)	0.0436** (0.0165)	0.0717** (0.0259)
SAT*online		0.0229 (0.0143)		−0.0470 (0.0336)
R^2	.1704	.1745	.0142	.0175
N	517		592	
Data Included	**720 < SAT ≤ 770**		**770 < SAT**	
Specification	**Baseline**	**SAT-online interaction**	**Baseline**	**SAT-online interaction**
Intercept	−3.8411 (11.3472)	9.3274 (17.0327)	2.0771 (21.5952)	27.0246 (31.9554)
Online	0.4654 (0.4204)	−23.2091 (22.8419)	0.6385 (0.4639)	−45.2755 (43.3566)
Math SAT score	0.0692*** (0.0151)	0.0517* (0.0227)	0.0606* (0.0273)	0.0290 (0.0404)
SAT*online		0.0315 (0.0304)		0.0581 (0.0548)
R^2	.0376	.0394	.0140	.0162
N	578		494	

Note: Standard errors are reported in parentheses underestimates. Baseline regressions first control for math SAT score alone, and then also interact math SAT score with instructional method. The Data Included field indicates the SAT score range of students included in the baseline and interaction specification.

***$p < .001$; **$p < .01$; *$p < .05$.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A2 Baseline credit-only probit results by total midterm score quartile.

Test score range/coefficient	Total score ≤ 121	122 ≤ Total score ≤ 134	135 ≤ Total score ≤ 143	144 ≤ Total score
Intercept	2.8255*** (0.7480)	7.6486* (3.0765)	−1.9485 (7.9680)	33.2860 (40.2223)
Midterm 1	−0.0122 (0.0076)	−0.0800*** (0.0238)	−0.0134 (0.0584)	−0.4066 (0.3199)
Midterm 2	−0.0257** (0.0080)	−0.0808** (0.0247)	0.0027 (0.0600)	0.0431 (0.2425)
Math SAT	−0.0013 (0.0010)	0.0024 (0.0016)	0.0012 (0.0027)	−0.0143 (0.0088)
N	297	310	297	305
Prop. CR	49.49%	18.71%	3.03%	0.66%

Note: Standard errors are reported in parentheses underestimates. The Test Score Range is the range of total midterm scores (Midterm 1 + Midterm 2) for students included in that column's results.
***$p < .001$; **$p < .01$; *$p < .05$.

results of students whose performance is improving across the midterms deciding to take the course for a grade based on their positive performance trajectory (Table A2).

Students in the upper half of the midterm score distribution are unlikely to take the course for credit-only. This is consistent with the results presented in Table 6. For these students there is no significant effect of either midterm score on the likelihood of taking the course for credit-only. This is not surprising given the generally low likelihood of these students taking the course for credit-only.

APPENDIX B: GRADING SCALES

Students' letter grade was determined through one of two methods—their overall total points earned across all assignments or their final exam grade. Grades based on overall points are determined by the relevant grading scale in columns 2 through 5 of Table B1. Letter grades from the final exam are determined by the grading scales in columns 6 through 9. Each student's grade was evaluated under both grading schemes. Students were then assigned the higher of the two letter grades. The largest percentage point change in the total points scale was an increase of 3 percentage points. The most relevant scale for this paper is the final exam scale. In percentage point terms, the largest change in this scale was an increase of 1.3 percentage points in the required grade to receive an A+.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE B1 Grading scales.

	Minimum total points score		Minimum overall % score		Minimum final exam points score		Minimum final exam % score	
Letter	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**
A+	384	384	96	96	192	175.2	96	97.3
A	354	366	88.5	91.5	178	160.8	89	89.3
A−	324	336	81	84	165	148.8	82.5	82.7
B+	312	318	78	79.5	162	144	81	80
B	303	310	75.75	77.5	159	141.6	79.5	78.7
B−	294	294	73.5	73.5	152	136.8	76	76
C+	278	278	69.5	69.5	148	132	74	73.3
C	264	262	66	65.5	137	122.4	68.5	68
C−	242	242	60.5	60.5	128	115.2	64	64
D+	234	234	58.5	58.5	124	110.4	62	61.3
D	230	230	57.5	57.5	120	108	60	60
D−	216	216	54	54	110	98.4	55	54.7

Note: Grading scales in 2019 and 2020. Minimum points scores show the lowest number of points to receive a letter grade. Minimum overall percent score shows the minimum grade percentage to receive a letter grade. Columns 2 through 5 show the grading scale for the semester grade. Columns 6 through 9 show the grading scale for the final exam.

23258012, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/soej.12635, Wiley Online Library on [18/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License