

Article

Enhancing Learning Outcomes in Econometrics: A 12-Year Study

Seyhan Erden 

Department of Economics, Columbia University, New York, NY 10027, USA; seyhan.erden@columbia.edu

Abstract: This paper presents the findings of 12 years of data from studying the teaching of econometrics. The first course on the topic of econometrics has always been a challenging course for both students and instructors. Students come from different quantitative backgrounds, and mostly with the prejudice that this is one of the most challenging courses in their academic career. We showed that using Classroom Response Systems (CRSs) such as polls closes the achievement gap between students from higher and lower quantitative levels. Besides students' performance, we also investigated instructor performance through teaching and course evaluations utilizing data from 38 classes over the course of 12 years. We showed that the instructor performance is higher under the in-class modality compared to the online modality and showed that this gap in performance between the two modalities widens as students' grades improve; a positive association between grades and instructor performance under the in-class modality exists; however, the association is negative under the online modality.

Keywords: teaching econometrics; teaching pedagogy

JEL Classification: A22; C13; C33; C80



Citation: Erden, S. Enhancing Learning Outcomes in Econometrics: A 12-Year Study. *Educ. Sci.* **2023**, *13*, 913. <https://doi.org/10.3390/educsci13090913>

Academic Editor: James Albright

Received: 9 July 2023

Revised: 17 August 2023

Accepted: 5 September 2023

Published: 8 September 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Econometrics, a cornerstone of Economics [1], has seen limited evolution in teaching methods despite data-driven changes [2]. Applied econometrics focuses on parameter estimation in economic theories [3]. Angrist and Pischke [4] revealed a lack of modern economic tools in undergraduate textbooks except for Stock and Watson's textbook. All the courses from which the data come, use the most recent editions of the Stock and Watson textbook [5].

Econometrics represents a pinnacle course within the field of Economics [1]. The examination of econometrics pedagogy was initially undertaken by the Econometrics Society in the early 1950s as an integral component of UNESCO's comprehensive investigation into social sciences instruction [6]. Angrist and Pischke surveyed econometrics textbooks published since the 1970s and compared them by the number of pages allocated to econometric topics. They suggest emphasizing causal examples of controlled statistical comparison over multivariate modeling and quasi-experimental methods.

Classroom Response Systems (CRSs) are becoming increasingly commonplace in educational settings [7]. As the frequency with which CRSs are used is increasing, it becomes increasingly important to define the affordances and limitations of these tools. Many researchers have evaluated the impact of CRSs in different classroom environments. In microbiology courses, it was found that using the CRS at the end of the class rather than at the beginning was more effective [8]. Wood and Shirazi used eight databases and identified six interconnected themes about audience response systems (ARS): engagement, interaction, anonymity, questioning instant feedback, and technological benefits limitation [9]. The themes reveal the complexity of student learning experiences (<https://www.sciencedirect.com/topics/social-sciences/learning-experience>) using ARS, which, when presented as

a model, contributes to current understanding and offers a framework of pedagogical conditions to consider when designing and implementing learning experiences (<https://www.sciencedirect.com/topics/social-sciences/learning-experience>) using ARS. Some studies found that the ability, motivation, and trigger capabilities of the Game-based Classroom Response System (GCRS) were the most important predictors of students' perception of learning [10].

This study shows that CRSs are associated with narrowing the gap between the achievement levels of students with different quantitative backgrounds when we control for the classroom, class time, class day, gender of the student, major, year, and several other effects. Obviously, students who are in more quant-heavy majors end up with higher grades in the econometrics course because it is a quant-heavy course. However, given the students' quantitative background, when polls are used in class as a CRS or ARS, the gap in the performance level of students with different degrees of quantitative backgrounds closes (Figure 1).

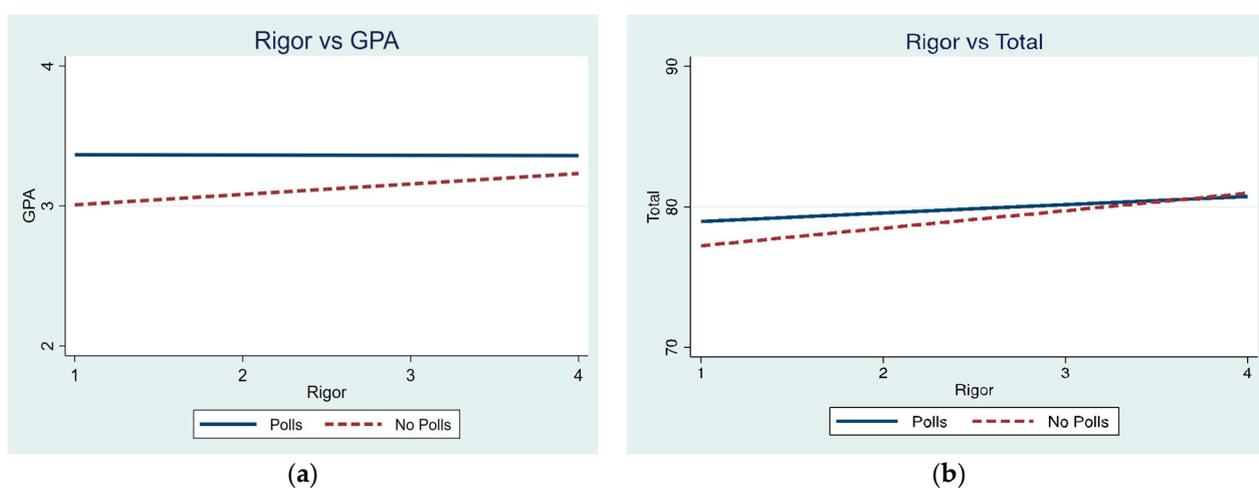


Figure 1. CRS use narrows the gap in GPA (the graph on the (a)) and total grade (on the (b)) of students coming from different academic quantitative rigor levels).

On the other hand, combining these results with teacher and course evaluations, some studies show that students' evaluations of teaching and learning are unrelated [11]. Earlier research supported the opposite. A large body of older international research confirmed a direct relationship between teacher quality/effectiveness and student learning [12–14]. We have shown that instructor performance, measured by student evaluations, depends on many factors. However, one common result is that, in general, instructor performance decreased during the pandemic under the online modality. The instructor performance is higher under the in-class compared to the online teaching modality. The gap in evaluations between the two modalities widens as students' grades improve; a positive association exists between students' grades and instructor performance in the in-class modality, but this association is negative under the online modality (Figure 2a). When the grades are curved (GPA), the association between grades and evaluations is insignificant under the in-class modality. Still, the same association is positive under online modality (Figure 2b). The instructor's performance, as measured by student evaluation, is negatively associated with class size. The higher the class size, the lower the evaluations. The evaluation gap under the two modalities narrows as class size becomes smaller; this means the negative association between evaluations and class size is stronger when teaching occurs in-class instead of online. These results have a couple of policy implications: (i) smaller class sizes (below 30 students) and (ii) the in-class modality works better for econometrics classes.

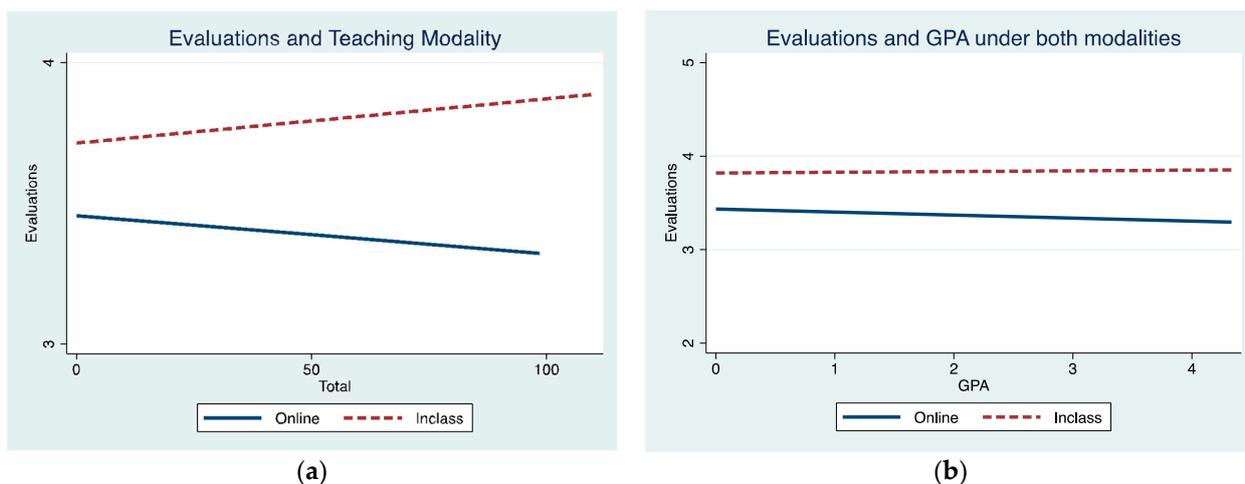


Figure 2. (a) Teaching evaluations are negatively correlated with students' performance measured using total grades under the online modality; however, they are positively correlated under the in-class modality. (b) GPA (curved grade) under in-class modality does not have a significant relation with teaching evaluation, and under online modality, the same negative correlation is observed.

2. Materials and Methods

Data are compiled from the undergraduate *Intro to Econometrics (UN3412)* course at Columbia University, Department of Economics, starting in Spring 2010 and ending in Fall 2022. Ten separate instructors taught this course in the given time; the data use one instructor who taught one to three sections every semester from Spring 2010 to Fall 2012 for 38 classes and 2678 students. Classroom Response Systems (*polls*) were implemented utilizing three different systems: clickers (Spring 2011 to Fall 2012; a total of four semesters, eight classes), Zoom polls (Spring 2020 and Summer 2021; for a total of five semesters including summer semesters; eight classes), and Poll Everywhere (Summer and Fall, 2022). Hence, CRSs were used in 18 out of 38 classes. Examples of Poll Everywhere can be found in Erden [15].

Variables and controls are listed in Table 1. There are three different dependent variables that measure the "outcome".

Table 1. Description of variables.

Variables/Controls	Description
Total	Total grade out of 100 in the course
GPA	The grade out of 4.0 in the econometrics course =1 (<i>Failing</i>) for $GPA < 2$;
Performance	=2 (<i>Average</i>) for $2 \leq GPA \leq 3.33$; =3 (<i>Above – average</i>) for $GPA > 3.33$
Final	Final exam grade
Midterm	Midterm exam grade
Psets	Problem Set average grade (total of 9 problem sets)
School	CC = Columbia College; EN = School of Applied Sciences and Engineering; GS = General Studies; BC = Barnard College
Level	U01 = Freshmen; U02 = Sophomores; U03 = Junior; U04 = Senior; U00 = Other
Affiliation	School and major combination of the student (ex: CCECON is Columbia College Economics Major, GSECMA is General Studies Econ-Math major)
Major	Declared major of the student (ex: Financial Engineering)
Econ	=1 if student majors in Economics only
Room	The physical classroom for the given semester
Polls	=1 if CRSs were used; = 0 otherwise

Table 1. Cont.

Variables/Controls	Description
Proctorio	=1 if this online proctoring platform was used for exams; = 0 otherwise
Male	=1 for male students; = 0 otherwise
Quant	A quantitative measure of mathematical rigor of the major. Majors characterized by substantial mathematical integration or a distinct mathematical emphasis, such as Mathematics, Statistics, Physics, or Econ-Math are assigned a score of 4. Majors lacking significant mathematical components, disciplines like Philosophy, Comparative Literature, or Art History, receive a score of 1. Majors intermittently employing mathematical principles are assigned a score of 2, while consistently integrating math without being predominantly math-oriented are assigned a score of 3
Polls × Quant	Interaction of Polls and Quant variables
Ins_Eval	Instructor evaluation by students (5 best, 1 worst)
Course_Eval	Course evaluations by students (5 best, 1 worst)

One is the “Total”, which is the total grade of students out of 100 calculated by a formula given on the syllabus. Basically, it takes 30% of the average of the problem set grades. There are nine problem sets, and the lowest grade is dropped. Then, between the midterm and final exams, the higher grade counts as 40%, and the lower one counts as 30%. Note that CRSs are used in some semesters but never count toward the course grade. They are used for the sole purpose of keeping students’ attention at the highest possible level. They are conducted after teaching a topic for 15 to 20 min. It is well-established that an average student’s attention span lasts about that time. Thus, when their attention span is fading, CRS is used with one or two questions about what has been taught. This method helps students to “wake up”. We have seen that this helps less successful students more. We suspect the reason for less successful students gaining more might be because those students’ attention spans might be shorter. Here, we assume students who are not interested in quantitative subjects would be less successful in quant-heavy topics or vice versa. However, with our data, we cannot test whether success and attention span simultaneously have causal effects on each other. Several scholarly works have been published on this issue [16].

The other dependent variable is GPA, the grade point average out of 4.0 in the econometrics class. It is calculated using the usual scale, where A is 4.0, A– is 3.67, B+ is 3.33, and so on. This grade is determined after the curve is applied, so class-fixed effects do not affect the results. The third dependent variable we used is performance. We divided students into three categories: the *above-average* category includes those students with A, A–, and B+ GPAs; the *average* category includes those students between C and B; and the *failing* category includes those students with C– and lower.

An essential variable of interest is called *Quant*, a measure of the student’s mathematical rigor. It is a numerical proxy for a given major’s mathematical (quantitative) rigor. Majors that are pure math or highly integrated with math—such as Mathematics, Statistics, Physics, or Econ-Math—are given a 4, whereas majors that involve no major mathematical component—such as Philosophy, Comparative Literature, or Art History—are given a 1. Majors that occasionally use math are given a 2, and majors that consistently, but not primarily, use math are given a 3. We introduce this variable as a proxy for students’ interests in quantitative subjects. Those students in more quantitative majors inherently have more interest in quant-heavy courses. Since econometrics is a quant-heavy course, students who show interest in quantitative majors tend to perform better in econometrics than those who show interest in less quantitative majors. However, we showed that in semesters in which we used polls, this gap in performance was closed (Figure 1).

Since student performance also depends on instructor performance, we used end-of-semester student evaluations to research instructor performance. The data use two of the questions from student evaluations. One is about how the instructor performed on a scale

of 1 through 5, and the other is about evaluating the overall course from 1 through 5; in both questions, 5 is excellent, 4 is very good, 3 is good, 2 is fair, and 1 is poor.

3. Results and Discussion

In general, CRSs are associated with higher GPAs. In the semesters CRSs used, GPAs are significantly higher when we control for all other variables. On average, CRSs are 9.3 percentage points higher grades and 0.66 higher GPA points at a 1% significance level—see Tables 2 and 3. In all three cases, CRSs narrow the gap between quantitatively less rigorous and more rigorous students (see Figure 1). We used the word “rigor” for quant-heavy (quantitatively rigorous) majors throughout this article. The horizontal axis is the students’ quantitative rigor level; the vertical axis is the students’ GPA (Figure 1a) or the total grade out of 100 (Figure 1b). Students at different quantitative rigor levels all benefit from CRSs by about the same amount. When CRSs are not used, quantitatively less rigorous students end up with lower GPAs than quantitatively more rigorous ones. As can be observed in Table 4, when polls are used in a particular class, the probability of failing students decreases by 6.7%, the probability of average students decreases by 17.8%, and the probability of *above-average* students increases by 24.5% compared to classes in which polls are not used. Definitions of *above-average*, *average*, and *failing* can be found in Table 1. CRSs narrow the gap between these students, helping the *average* and *failing* students most and helping the *above-average* students less—see Tables 4 and 5.

Table 2. Total grade out of 100.

	(1)	(2)	(3)	(4)
Variables	Total	Total	Total	Total
Polls	2.783 ** (1.183)	4.810 ** (1.974)	5.069 ** (1.995)	9.271 *** (3.281)
Proctorio	−5.536 *** (1.678)	−0.130 (2.552)	−0.216 (2.559)	−0.143 (2.557)
Tues-Thur (vs. MW)		1.491 ** (0.744)	1.499 ** (0.740)	1.641 ** (0.746)
Male		−1.093 * (0.652)	−1.351 ** (0.653)	−1.330 ** (0.659)
Econ			−3.436 *** (0.778)	−2.872 *** (0.866)
Quant				1.059 ** (0.502)
Polls × Quant				−2.019 * (1.106)
Constant	78.99 *** (0.354)	80.49 *** (4.471)	80.74 *** (4.478)	81.04 *** (4.674)
Fixed Effects:				
Time of the day effects	No	Yes	Yes	Yes
Day of the week effects	No	Yes	Yes	Yes
Classroom effects (including online)	No	Yes	Yes	Yes
School effects	No	Yes	Yes	Yes
Quant and polls interactions	No	No	No	Yes
Observations	2316	2311	2311	2277
R-squared	0.003	0.099	0.107	0.112

Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3. Grade point average (for the course).

	(1)	(2)	(3)	(4)
Variables	GPA	GPA	GPA	GPA
Polls	0.232 *** (0.0637)	0.331 *** (0.105)	0.352 *** (0.107)	0.658 *** (0.188)
Proctorio	−0.0998 (0.0948)	−0.197 * (0.119)	−0.204 * (0.120)	−0.211 * (0.120)
Male		−0.0447 (0.0348)	−0.0591 * (0.0348)	−0.0582 * (0.0352)
Econ			−0.211 *** (0.0418)	−0.170 *** (0.0461)
Quant				0.0565 ** (0.0269)
Polls × Quant				−0.126 ** (0.0627)
Constant	3.107 *** (0.0182)	3.351 *** (0.217)	3.365 *** (0.216)	3.324 *** (0.230)
Fixed Effects:				
Time of the day effects	No	Yes	Yes	Yes
Day of the week effects	No	Yes	Yes	Yes
Classroom effects (including online)	No	Yes	Yes	Yes
School effects	No	Yes	Yes	Yes
Quant and Polls interactions	No	No	No	Yes
Observations	2204	2201	2201	2171
R-squared	0.005	0.085	0.096	0.099

Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4. Change in performances (ordered logit).

	(1)
Ordered Logit	With Polls Probabilities Change
Probability (Perf = 1) <i>failing students</i>	↓ by 6.7%
Probability (Perf = 2) <i>average students</i>	↓ by 17.8%
Probability (Perf = 3) <i>above-average students</i>	↑ by 24.5%

↓ decreases, ↑ increases.

Table 5. Performance (ordered logit).

	(1)	(2)	(3)	(4)
Variables	Performance	Performance	Performance	Performance
Polls	0.653 *** (0.201)	1.026 *** (0.295)	1.080 *** (0.303)	1.704 *** (0.564)
Proctorio	−0.557 * (0.284)	−0.706 * (0.384)	−0.725 * (0.389)	−0.687 * (0.386)
Tues-Thur (vs. MW)		0.277 *** (0.103)	0.280 *** (0.104)	0.283 *** (0.104)
Econ			−0.334 *** (0.114)	−0.339 *** (0.117)
Quant			0.131 ** (0.0648)	0.151 ** (0.0683)
2.Quant#c.Polls				−0.509 (0.535)
3.Quant#c.Polls				−1.076 ** (0.546)

Table 5. Cont.

	(1)	(2)	(3)	(4)
Variables	Performance	Performance	Performance	Performance
4.Quant#c.Polls				−0.210 (0.786)
/cut1	−2.639 *** (0.0855)	−3.294 *** (0.567)	−3.308 *** (0.591)	−3.269 *** (0.593)
/cut2	−0.292 *** (0.0440)	−0.832 (0.559)	−0.838 (0.583)	−0.796 (0.585)
Fixed Effects:				
Gender dummy	No	Yes	Yes	Yes
Time of the day effects	No	Yes	Yes	Yes
Day of the week effects	No	Yes	Yes	Yes
Classroom effects (including online)	No	Yes	Yes	Yes
School effects	No	Yes	Yes	Yes
Observations	2316	2311	2277	2277

Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The two most important results that contribute to the education of econometrics literature are as follows: (1) CRSs not only improve all students' performances in class but also close the gap in the performances of students with different quantitative backgrounds. (2) teaching evaluations not only decrease during the COVID-19 era but also the gap in evaluations widens as students' grades improve; a positive association exists between students' grades and instructor performance in the in-class modality, but this association is negative under the online modality. The instructor's performance, as measured by student evaluation, is negatively associated with the class size. The higher the class size, the lower the evaluations; this negative association is less strong under the online modality.

Some of the other expected but still exciting results are as follows:

- Proctorio (an online proctoring platform) has a significantly negative association with GPA (0.21 less GPA points). However, since Proctorio was only used when the class was taught online, the significant negative effect of Proctorio on grades might have captured some adverse effects of online learning. However, as for the total grades, Proctorio's association with total grades is not significant. The p -value for this coefficient is 0.117; this high p -value might be because of the high correlation between Proctorio and Polls (correlation is 0.41—a multicollinearity problem). The higher the correlation, the more likely the individual significance test is to under-reject. Here, under-rejection refers to the fact that standard errors of the sample slope increase as the correlation between regressors (independent variables in the regression) rises; hence, the test statistic (z -score) decreases because the standard error of the sample slope is in the denominator. Thus, we are less likely to reject a significance test for a given critical value, such as 1.64, 1.96, or 2.58; this is what we refer to as under-rejection here. We are more likely to find the population slope insignificant as the correlation between regressors increases. As the correlation approaches one, the variance of each coefficient approaches infinity, as does their standard error, which causes higher t -stats and under-rejection of the null hypothesis. The joint significance test is a commonly used "solution" in the literature. Due to the joint significance of these two variables, we decided to keep the Proctorio variable in the regression to avoid omitted variable bias—Tables 2 and 3.
- Comparison of different schools within Columbia: we noticed that in terms of performance in the econometrics class, the order is as follows: Engineering, Columbia College, General Studies, followed by the category "others" (exchange students and professional studies), and those from Barnard College, which is a liberal arts college, and shows comparatively lower performance due to the quantitative nature of

the course. Notably, Engineering, Financial Engineering, and Electrical Engineering majors outperform Operations Research—Tables 2 and 3.

- Regarding academic performance, classes held on Tues-Thurs exhibit a slight advantage over those scheduled on Mon-Wed. This phenomenon remains consistent when analyzed using an ordered logit regression—Table 4.
- Students who take the class in the Fall semester perform better than those taking the class in the Spring semester (Tables 2–4).
- Any time the class was taught in one particular classroom, students received significantly higher GPA points (*p*-value of 0.02) compared to the semesters when the class was taught in other classrooms. This difference may not be a causal effect because this particular classroom is magnificent and used in many Hollywood movies; it may be giving students extra motivation.
- Morning classes (10:10 am) display superior outcomes compared to post-lunch ones (1:10 pm); with the subsequent highest performance observed in the afternoon (2:40 pm and 4:10 pm)—Tables 2–4. However, it is entirely possible that students with certain schedules and habits select morning classes, so a causal interpretation may not be possible.
- Among the academic classification, sophomores demonstrate the highest performance, followed by juniors and seniors. However, the interaction with CRSs is implemented, juniors perform significantly better—Tables 2–4.
- Students pursuing pure Economics majors (excluding double majors or concentrations such as Econ-Math, Econ-Stats, Econ-Political Science, or Financial Economics) exhibit a notable performance decline in the class—Tables 2–4.

On the other hand, under the online modality, teachers’ evaluations have decreased significantly. Figure 2 shows the relationship between teachers’ evaluations and class size in the online and in-class modalities. Analyzing Table 6, we can easily see the positive associations of polls and course evaluation and negative associations of online modality, Proctorio, and class size with teacher evaluations. There are some exciting but expected results from the regressions in Table 6. They are as follows:

- The gap in evaluations between online and in-class narrows slightly as students’ performances, measured by total grade out of 100, increase.
- Class size, COVID-19/online teaching, and the use of Proctorio are associated with lower evaluation results.
- CRS (polls) and students’ performance, measured as students’ total grade out of 100 in a class, are associated with higher evaluation results.

Table 6. Teachers’ evaluations with Total grade out of 100.

	(1)	(2)	(3)	(4)
Variables	Ins_Eval	Ins_Eval	Ins_Eval	Ins_Eval
Course_Eval	0.533 *** (0.0118)	0.591 *** (0.0209)	0.591 *** (0.0209)	0.592 *** (0.0210)
Class_size	−0.00395 *** (0.000238)	0.00226 *** (0.000387)	−0.00226 *** (0.000387)	−0.00236 *** (0.000445)
Total	0.00127 *** (0.000278)	0.00213 *** (0.000244)	0.00214 *** (0.000245)	0.00236 *** (0.000274)
Online/COVID−19	−0.247 *** (0.00968)	−0.175 *** (0.0202)	−0.175 *** (0.0202)	−0.145 *** (0.0451)
Polls	0.0580 *** (0.00542)	0.0420 *** (0.0150)	0.0423 *** (0.0150)	−0.0108 (0.0223)
Proctorio	−0.0756 *** (0.0126)	−0.0327 (0.0294)	−0.0326 (0.0294)	0.00597 (0.0296)
Tues-Thur (vs. MW)		0.0306 *** (0.00967)	0.0305 *** (0.00968)	0.0198 * (0.0114)

Table 6. Cont.

	(1)	(2)	(3)	(4)
Variables	Ins_Eval	Ins_Eval	Ins_Eval	Ins_Eval
Online × Total				−0.00204 *** (0.000292)
Online × Class_size				0.00182 *** (0.000545)
Constant	2.120 *** (0.0571)	1.808 *** (0.110)	1.804 *** (0.109)	1.791 *** (0.111)
Fixed Effects:				
Time of the day effects	No	Yes	Yes	Yes
Day of the week effects	No	Yes	Yes	Yes
Classroom effects	No	Yes	Yes	Yes
School effects	No	Yes	Yes	Yes
Online—Total grade interactions	No	No	No	Yes
Observations	2696	2688	2688	2633
R-squared	0.557	0.696	0.696	0.695

Robust standard errors in parentheses: *** $p < 0.01$, * $p < 0.1$.

There are some policy implications of the results of this research as well: econometrics should not be taught in large sections. There might be an optimum number of students for these classes, but we will study that in future research. For now, we observe that classes with sizes below 30 produce more satisfying learning experiences than those 86 and above.

Using Technology in the Classroom

This section presents a first-hand account of the transition from in-person, online, and back to in-person teaching, outlining the associated challenges and opportunities. Prior to the COVID-19 pandemic in Spring 2020, conventional teaching tools and methods were predominant among instructors [15]. However, the shift to online instruction necessitated the swift adoption of new tools and techniques to ensure adequate student learning, presenting a global pedagogical challenge. Upon the resumption of in-person instruction at Columbia University in Fall 2021, certain technologies utilized during the online period were seamlessly integrated.

The following effective technologies are highlighted here, hoping that they would be helpful for instructors in the profession:

- For the shift to online teaching, replacing the traditional blackboard with a tablet device and leveraging a software application as a digital “blackboard” has been highly effective.
- Initially, a combination of PowerPoint slides and the tablet’s blackboard-like software was employed. Later, this evolved into converting slides into pdf files, with added spacing between slides.
- This modification allowed real-time annotations, enabling simultaneous viewing of slides and annotations, surpassing the effectiveness of traditional blackboard.
- Upon returning to in-person instruction, the tablet screen was displaced on a larger interface, serving as an efficient alternative to “chalk and board” teaching. This transition offered advantages such as maintaining face-to-face interactions while writing and avoiding the need to turn away from students.
- Furthermore, the display screen provided a more comprehensive visual interface than the conventional blackboard.
- The use of Zoom Polls was replaced by Poll Everywhere. Students use their cell phones to answer poll questions in the in-class teaching modality.

This section offers insights into a pedagogical journey marked by innovative adaptations, effective technology integration, and the seamless continuation of successful online tools in the physical classroom.

4. Conclusions

Utilizing 12 years of data from “Intro to Econometrics”, a course offered by Columbia University’s Department of Economics, we assess the use of CRSs and teaching evaluations. Our study reveals noteworthy insights with potential implications for instructors of analogous courses across diverse universities.

The most significant finding is about CRSs; examples can be found in [15]. When we use CRSs, students’ performance measured by three different variables (*GPA*, *Total*, *Performance*) improved. When CRSs were used for online classes, the GPAs of students from different quantitative rigor levels were flat, meaning that CRSs were helpful to all students from different quantitative backgrounds. When we did not use CRS, students from quantitatively heavier majors performed better. We interpret this as CRSs narrowing the gap in students’ performances from different quantitative levels (Figure 1a). When we measure students’ performance level using a grade out of 100, recall that this is not the curved grade; as students’ majors’ math content intensifies, performance in class becomes better under both modalities (Figure 1b). However, this difference is more visible for the semesters we do not use CRSs. The regression line is steeper for CRS semesters compared to no-CRS semesters. Both results in Figure 1 are associated with the help of CRS, improving overall grades and narrowing the gap between students with different degrees of quantitative rigor levels.

Using ordered logit regression for different performance levels, the *above-average*, the *average*, and the *failing*, we have shown that the probability of the *above-average* increases and the probability of the other two decreases for a given student with the use of CRSs, further proving the narrowing in the performance gap between students at different rigor levels (here the term “rigor”, refers to the math level of the student’s major). The second most significant finding is about instructor performance, as measured by teaching evaluations; under the online modality, evaluations decrease, and the gap in evaluations between the two modalities widens with better student performance while the same gap narrows with class size. Given the in-class modality, instructor evaluations and students’ total grades out of 100 are positively associated. However, under the online modality, this association is negative (Figure 2a). On the other hand, under both modalities, instructor performance is negatively associated with class size. This negative association is steeper (stronger) for the in-class modality than for the online modality (Table 6, column 4). This result has a policy implication: the class size for econometrics classes should be capped at the lowest feasible number of students per section, and the preferred modality is in-class for econometrics teaching.

There are many control variables used in this research. We have observed some interesting but expected results for some control variables. Proctorio substantially reduced performance. Engineering students excelled, aligning with the quantitative nature of econometrics. A “Monday effect” emerged, with Tues-Thurs classes exhibiting slightly but significantly higher performance than their Mon-Wed counterparts. Students in the Fall semesters outperformed those in the Spring semesters. The “senioritis effect” was evident, with seniors exhibiting notably lower performance in the course.

In conclusion, this study enhances our understanding of effective teaching methods for the challenging course econometrics. Analyzing extensive data underscores the efficacy of CRSs in facilitating student learning and bridging achievement gaps. The study also gives insights into effective learning of econometrics by analyzing the instructor performance measured by student evaluations. These findings offer usable insights for instructors, enhancing their effectiveness in teaching econometrics and benefiting the broader field. Moreover, the research carries policy implications.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable; no animals were involved in this study.

Informed Consent Statement: Not applicable; no human patients were involved in this study.

Data Availability Statement: Data is not public due to the privacy of students' grades.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Klein, C.C. Econometrics as a Capstone Course in Economics. *J. Econ. Educ.* **2013**, *44*, 268–276. [CrossRef]
2. Asarta, C.J.; Chambers, R.G.; Harter, C. Teaching Methods in Undergraduate Introductory Economics Courses: Results From a Sixth National Quinquennial Survey. *Am. Econ.* **2020**, *66*, 18–28. [CrossRef]
3. Johnson, B.K.; Perry, J.J.; Petkus, M. The Status of Econometrics in the Economics Major: A Survey. *J. Econ. Educ.* **2012**, *43*, 315–324. [CrossRef]
4. Angrist, J.D.; Pischke, J.-S. Undergraduate Econometrics Instruction: Through Our Classes, Darkly. *J. Econ. Perspect.* **2017**, *31*, 125–144. [CrossRef]
5. Tintner, G. The Teaching of Econometrics. *Econometrica* **1954**, *22*, 77. [CrossRef]
6. Fies, C.; Marshall, J. Classroom Response Systems: A Review of the Literature. *J. Sci. Educ. Technol.* **2006**, *15*, 101–109. [CrossRef]
7. Suchman, E.; Uchiyama, K.; Smith, R.; Bender, K. Evaluating the Impact of a Classroom Response System in a Microbiology Course. *Microbiol. Educ.* **2006**, *7*, 3–11. [CrossRef] [PubMed]
8. Wood, R.; Shirazi, S. A systematic review of audience response systems for teaching and learning in higher education: The student experience. *Comput. Educ.* **2020**, *153*, 103896. [CrossRef]
9. Ashtari, S.; Taylor, J. Winning together: Using game-based response systems to boot perception of learning. *Int. J. Educ. Dev. Using Inf. Commun. Technol.* **2021**, *17*, 123–141.
10. Stock, J.H.; Watson, M.W. *Introduction to Econometrics*, 4th ed.; Addison Wesley: Boston, MA, USA, 2019.
11. Uttl BWhite, C.A.; Gonzalez, D.W. Meta-analysis of faculty's teaching effectiveness. *Stud. Educ. Eval.* **2017**, *54*, 22–42.
12. Goldhaber, D.; Anthony, E. Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching. *Rev. Econ. Stat.* **2007**, *89*, 134–150. [CrossRef]
13. Hattie, J.A.C. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*; Routledge: London, UK, 2009.
14. Sanders, W.; Ashton, J.; Wright, S. *Comparison of the Effects of NBPTS-Certified Teachers with Other Teachers on the Rate of Student Academic Progress*; U.S. Department of Education and National Science Foundation: Washington, DC, USA, 2005.
15. Erden, S. *Leverage Educational Technology to Create Engaging Classroom Experiences*; Columbia University: New York, NY, USA, 2023. Available online: <https://ctl.columbia.edu/transformations/faculty/seyhan-erden/> (accessed on 8 June 2023).
16. Oberauer, K. Working Memory and Attention—A Conceptual Analysis and Review. *J. Cogn.* **2019**, *2*, 36. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.