Economics Bulletin

Volume 39, Issue 4

Student engagement and larger class enrollments: evidence from a growing mid-sized university

Michael Gove
University of North Georgia

Abstract

With increasing enrollment and class sizes in many colleges and universities, the potential connection between class size and student engagement has growing relevance for students, faculty, and administrators alike. I examine this potential connection, focusing empirically on a sample that allows for capturing how class size is related to various engagement measures on student evaluations of instructors over a nine-semester period in a College of Business at a growing mid-sized university. Across multiple specifications varying in both functional form and inclusion of fixed effects controlling for differences in instructors and courses, I find a consistently significant negative relationship between the class size and instructor evaluation ratings. This negative relationship is largest when increasing class size at lower enrollment levels, the negative effect diminishing somewhat in size as class sizes increase.

This work was supported by the UNG Mike Cottrell College of Business Research Grant.
Citation: Michael Gove, (2019) "Student engagement and larger class enrollments: evidence from a growing mid-sized university", *Economics Bulletin*, Volume 39, Issue 4, pages 2550-2565
Contact: Michael Gove - michael.gove@ung.edu.
Submitted: February 12, 2019. **Published:** November 12, 2019.

1. Introduction

Given recent trends of increasing enrollment and class sizes in many colleges and universities, the potential connection between class size and student engagement finds growing relevance in the context of higher education, whether examined from the perspective of students, faculty, or administrators. Students care about this connection particularly if student engagement is correlated with actual achievement and learning (see Carini et al. (2006) for examination of this topic); faculty care because student outcomes may be affected by class size, and measures of engagement (specifically, teaching evaluations) are of utmost importance in most promotion and tenure processes; finally, administrators care for the same reasons students and faculty pay attention to this relationship, as well as for the fact that perceptions of class size, student engagement, and outcomes clearly affect overall perceptions of the quality of an entire academic program, potentially subsequently affecting future enrollment, funding, and growth of a program.

I examine the relationship between class size and selected questions answered by students on teaching evaluations, allowing for one measure of the connection between class size and student engagement.[1] While this relationship has been previously examined in the literature, many earlier studies ignore the reality that differences across instructors and courses can contribute to evaluation and engagement outcomes (McConnell and Sosin (1984), DeCanio (1986), and Siegfried and Kennedy (1995), among others). More recent studies connecting class size with a variety of student outcomes include a focus on academic interactions as the outcome variable (Beattie and Thiele, 2016), total student load as a potential explanatory variable (Monks and Schmidt, 2011), and estimation strategies allowing for claims of measuring the causal effect of class size on outcomes (Shin, 2011, and Sapelli and Illanes, 2016). My data importantly allow for the inclusion of instructor, course, and instructor-course fixed effects, controlling for this heterogeneity that would otherwise bias any estimates, and providing one important difference from more recent studies such as Bedard and Kuhn (2008) and Ragan and Walia (2010). Furthermore, focusing on the case of the Mike Cottrell College of Business (MCCB) at the University of North Georgia provides an interesting difference from these recent studies, as the MCCB generally has smaller class sizes than many comparable university programs. These characteristics of the data allow for: (1) exploiting variation in class size across parallel sections of the same course taught by the same instructor in the same semester, and (2) exploring whether a significant relationship exists for variation in class sizes at smaller average initial class size levels, rather than focusing on the largest class sizes of hundreds of students, increasingly found in many universities. My findings point to the existence of a small yet significant negative relationship between class size and average instructor ratings on teaching evaluations, consistent across varied specifications and robustness checks. Specifically, given the preferred specification with class size indicators grouping enrollment by tens and instructor, course, and instructor-course fixed effects included, increasing class size from 0-9 to 10-19 is associated with an average decrease of

[1] The selected questions from the teaching evaluations provide only one measure, albeit clearly imperfect, of student engagment. The intent of this paper is not to enter into the debate over how well evaluations capture engagement, however recognizing this shortcoming gives one reason for performing the robustness checks found in Section 4.

0.07 for instructor ratings. This decrease increases slightly to 0.11, followed by decreases of 0.06, 0.09, 0.05, and 0.09 points, respectively, as class size is increased over the next six subsequent enrollment groups of ten, with the final decrease of 0.03 points associated with an increase in class size from 60-69 to greater than 70. In continuation, Section 2 describes the data, Section 3 details the fixed effects model, I discuss the results, varied specifications, and robustness checks in Section 4, while Section 5 concludes.

2. Data and Description

I use teaching evaluation data from the *COB* across nine semesters to carry out my empirical investigation. The data from 2014 to 2016 provides 1711 observations, consisting of 134 unique courses taught by 152 different instructors during the three-year period. Observed information useful for estimation includes class size, academic department, instructor, course, semester, year, and evaluation ratings. Each evaluation rating used in estimation is an aggregate of observed evaluation ratings, first summed and averaged over all individual student ratings for each given course section in response to one statement on the university's teaching evaluations.

$R_{ysci} = \frac{\sum_{j=1}^{N_{ysci}} r_{yscij}}{N_{ysci}}$, (1)

where R is the aggregate rating for year y, semester s, course c, and instructor i; N is the number of student responses for r in each given course section. This aggregate rating is then summed with and averaged across three additional aggregate ratings, the four statements from the teaching evaluations selected given their relevance to student engagement.

$Z_{ysci} = \frac{R_{ysci,S1}+R_{ysci,S2}+R_{ysci,S3}+R_{ysci,S4}}{4}$, (2)

where Z is the composite rating of interest for estimation, and $S1$, $S2$, $S3$, and $S4$ indicate the four selected teaching evaluation statements, respectively. [2] Ratings are in response to the following four statements: (S1) Thought provoking ideas and concepts were introduced, (S2) The instructor encouraged student involvement/questions, (S3) The instructor challenged me to think critically, and (S4) Overall the course was effective in helping me learn. Student answers range from 1 to 5, corresponding to Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, and Strongly Agree, respectively. Table 1 provides summary statistics, focusing on composite ratings and class size across the nine available semesters, while Figure 1 includes all average instructor ratings from each course examined related to class size.

[2]While this strategy is similar to that of Bedard and Kuhn (2008), the aggregated rating in the measurement of the dependent variable under examination provides an important difference. Furthermore, in Section 4, varied evaluation statements are used in estimations as checks for robustness, assuring that results are not simply reliant on statement selection.

Table 1. Summary Statistics					
	Instructor ratings Z_{tci}		Enrollment		
Semester	Mean (s.d.)	Min/max	Mean (s.d.)	Min/max	Observations
Fall 2016	4.14 (0.63)	1.6/5	34.43 (13.10)	3/74	268
Summer 2016	4.29 (0.50)	3.1/5	23.54 (11.54)	1/44	89
Spring 2016	4.12 (0.58)	2.3/5	35.24 (14.79)	2/86	254
Fall 2015	4.07 (0.64)	1.3/5	34.95 (13.12)	1/79	246
Summer 2015	4.16 (0.89)	2/5	25.20 (11.94)	1/44	89
Spring 2015	4.10 (0.61)	2.3/5	35.03 (12.18)	3/77	226
Fall 2014	4.08 (0.60)	2.4/5	34.76 (11.70)	2/75	230
Summer 2014	4.11 (0.73)	1/5	22.84 (10.97)	1/47	81
Spring 2014	4.14 (0.56)	1.9/5	31.77 (12.39)	1/75	228
Total	4.12 (0.63)	1/5	32.80 (13.32)	1/86	1711

Figure 1: Class size and instructor ratings (n=1711)

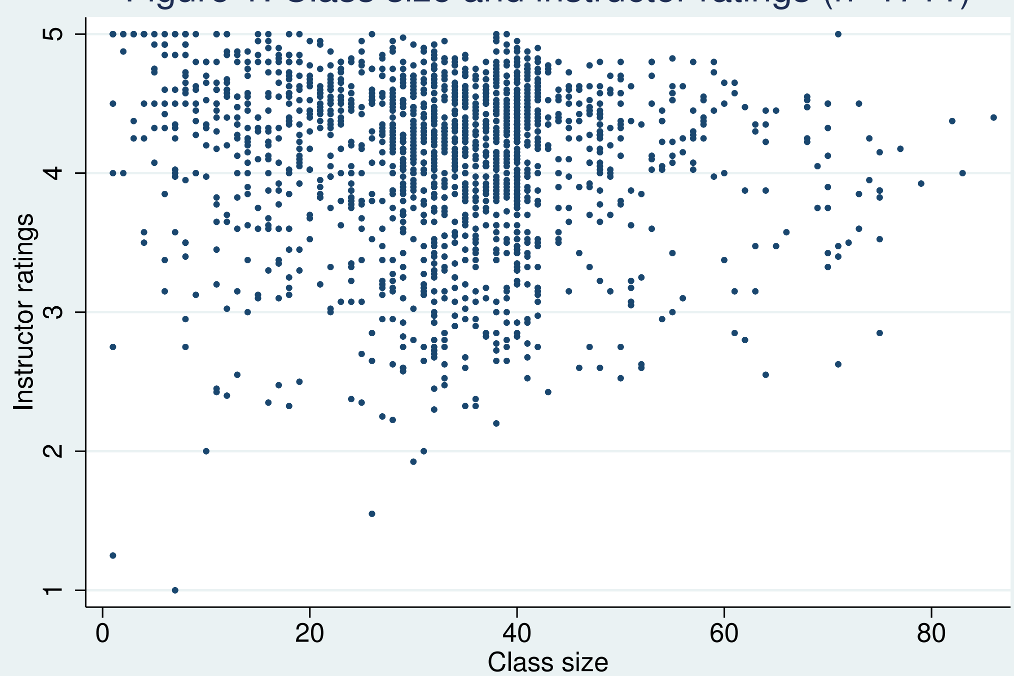


Table 2 highlights average composite ratings categorized by class size, as well as test statistics for the null hypothesis that any given mean is the same as the previous category's mean. Large, significant drops in instructor ratings are reflected both when class enrollment increases from less than 20 to between 20 and 39 students and when enrollment increases from between 40 and 59 to more than 60 students. Perhaps surprisingly, average ratings actually increase when course enrollment jumps from between 20 and 39 to between 40 and 59 students, although this difference in means is not statistically significant. When course enrollments are categorized by tens rather than twenties, a similar trend emerges although the significance of the differences in means across categories diminishes. Average ratings steadily decrease as class enrollment increases, essentially leveling out over 40-49 and 50-59 students, then decreasing again at 60 and above.

Table 2. Enrollment size and average instructor ratings Z_{tci}

Class size	Class means (1)	Means t-stat (2)	Sample size (3)	Class size	Class means (4)	Means t-stat (5)	Sample size (6)
<20	4.22 (0.81)	-	285	<10	4.27 (0.99)	-	90
20-39	4.10 (0.58)	2.70	963	10-19	4.19 (0.70)	0.79	195
40-59	4.13 (0.57)	-0.85	403	20-29	4.12 (0.60)	1.15	293
60+	3.94 (0.56)	2.39	60	30-39	4.09 (0.58)	0.79	670
				40-49	4.13 (0.57)	-1.14	328
				50-59	4.11 (0.60)	0.35	75
				60-69	3.98 (0.61)	0.99	31
				70+	3.90 (0.51)	0.59	29

Parentheses indicate standard deviations. T-statistics result from t-tests for the null hypothesis that a mean for a particular class size category is the same as the mean for the class size category immediately smaller.

3. Fixed Effects Model

While Table 2 gives a worthwhile first glimpse at the relationship between class size and instructor ratings, in order to truly capture this relationship, other variables that may contribute to the variation in ratings must be controlled for. Perhaps most importantly, variation of instructors and courses could have an effect on average instructor ratings; differences across instructors regarding grading difficulty and teaching ability and differences in difficulty of course material all potentially affect instructor ratings. If this variation is not accounted for, any estimation of the effect of class size on average ratings would clearly result in biased estimates. Given that all instructor and course information is available for the surveyed semesters, I follow Bedard and Kuhn (2008) and Ragan and Walia (2010) and rely on the following simple fixed effects model in Equation 1, additionally controlling for variation in semester in the academic calendar and academic department.

(Eq. 1) $Z_{tci} = \alpha_i + \beta f\left(X_{tci}\right) + \gamma Y_{tci} + \varepsilon_{tci}$,

where Z_{tci} represents the average instructor rating defined in Section 2 for time period t, course c, and instructor i, β is a vector of instructor fixed effects interacting with some function f of class size X_{tci}, Y_{tci} captures time-varying variables including term in academic calendar and academic department, and ε is the associated error term.[3] The use of fixed effects in lieu of random effects more likely controls for any omitted variable bias, and facilitates my objective of controlling for time-invariant variables by partialling them out and

[3]In order to minimize any worries of potential heteroskedasticity, all models are weighted by the square root of the number of student evaluations in each class.

not focusing on their estimated effects (i.e., my focus is on the class size effect on outcomes, not on any given single instructor's or course's effect on outcomes). As discussed in Bedard and Kuhn (2008), use of the fixed effects model provides three advantages compared to much of the previous relevant literature: (1) inclusion of instructor fixed effects permits addressing differences across instructors, (2) large sample size allows for a variety of functional forms regarding $f\left(X_{tci}\right)$, and (3) course fixed effects are easily introduced as an additional check, given that general differences in courses may be related to instructor ratings.

In turn, as an additional check I estimate a second fixed effects model, now including course fixed effects alongside instructor fixed effects. Variability in the difficulty of material or in the level of student interest across different courses may also contribute to differences in average instructor ratings; Equation 2 reflects this additional control.

(Eq. 2) $Z_{tci} = \alpha_i + \beta f\left(X_{tci}\right) + \delta_c + \gamma Y_{tci} + \varepsilon_{tci}$,

where δ_c is a vector of course fixed effects.

Finally, in the most demanding specification, I agument Equation 2 to include instructor-course fixed effects, allowing for exploiting only the variation in class size across parallel sections of the same course taught by the same instructor in the same semester. Equation 3 expresses this ideal specification.

(Eq. 3) $Z_{tci} = \alpha_i + \beta f\left(X_{tci}\right) + \delta_c + \tau_{ci} + \gamma Y_{tci} + \varepsilon_{tci}$,

where τ_{ci} is a vector of instructor-course fixed effects.

4. Results and Discussion

Although including instructor, course, and instructor-course fixed effects forms the preferred specification for the reasons already described, I start with the cross-section model results as a benchmark for comparison, focusing on the flexible set of class size indicators grouping class size by tens.[4] Column 3 from Table A in the Appendix reflects a relationship between class size and instructor ratings that starts negative, evaluation ratings dropping by 0.08, 0.06, and 0.01 points as class size increases from 0-9 to 10-19, 10-19 to 20-29, and 20-29 to 30-39, respectively. However, this trend reverses when class size increases from 30-39 to 40-49, ratings increasing 0.05 points before falling 0.03 points with a class size increase from 40-49 to 50-59. Finally, the negative relationship prevails with class size moving from 50-59 to 60-69, corresponding to the largest and most significant ratings drop of 0.18 points before turning slightly positive for class sizes above 70.

Table 3 results give a more complete look at the actual relationship between class size and instructor ratings, given the added fixed effects. Results from Column 2 control for instructor differences, revealing a consistently negative and statistically significant relationship between class size and evaluation ratings. Increasing class size from 0-9 to 10-19 is associated with a ratings decrease of 0.13 points, although this movement is not statistically significant. A small drop of 0.03 points is followed by decreases consistent in size of 0.07, 0.07, 0.07, and 0.06 points as class size moves from 10-19 to 20-29, 20-29 to 30-39, 30-39 to 40-49, 40-49 to 50-59, and 50-59 to 60-69, respectively. Notably, all of these decreases are highly statistically

[4] Results from quadratic specifications and class size indicators grouping class size by twenties are also highlighted, however class size grouped by tens is preferred given its precision.

significant. Estimates in Column 4 reflect adding fixed effects for course differences to the instructor fixed effects. This addition results in the ratings values being generally higher, however the negative relationship turns more negative for smaller class sizes than before, with a ratings decrease of 0.16 points and 0.08 points when class sizes increase from 0-9 to 10-19 and 10-19 to 20-29, respectively. For increases over the remaining larger class size categories, results nearly match those from Column 2. Finally, estimates in Column 6 reflect the addition of instructor-course fixed effects. Ratings values are generally slightly lower than the other two specifications, with an initial smaller decrease of 0.07 points when increasing class size from 0-9 to 10-19. Decreases of 0.11, 0.06, 0.09, 0.05, and 0.09, respectively, follow moving across the next six class size categories.[5]

[5] Results associated with grouping class size by twenties are highlighted in Table B of the Appendix.

Table 3: Impact of enrollment size on average instructor ratings, Z_{tci}						
	Instructor fixed effects		Instructor and course fixed effects		Instructor, course, and instructor-course fixed effects	
	(1)	(2)	(3)	(4)	(5)	(6)
Class size	**-0.0091 (0.0031)**		**-0.0137 (0.0036)**		**-0.0127 (0.0036)**	
Class size $^2/100$	0.0027 (0.0038)		0.0076 (0.0043)		0.0060 (0.0042)	
Class size 10-19		-0.1289 (0.0804)		-0.1620 (0.0901)		-0.0657 (0.0915)
Class size 20-29		**-0.1613 (0.0770)**		**-0.2420 (0.0908)**		-0.1723 (0.0910)
Class size 30-39		**-0.2311 (0.0743)**		**-0.3055 (0.0900)**		**-0.2337 (0.0908)**
Class size 40-49		**-0.3053 (0.0766)**		**-0.3833 (0.0925)**		**-0.3190 (0.0932)**
Class size 50-59		**-0.3760 (0.0855)**		**-0.4521 (0.1012)**		**-0.3735 (0.1022)**
Class size 60-69		**-0.4358 (0.0980)**		**-0.5074 (0.1128)**		**-0.4669 (0.1131)**
Class size 70+		**-0.4946 (0.1050)**		**-0.5507 (0.1183)**		**-0.4932 (0.1171)**
F-statistic	12.64	12.20	8.19	8.00	7.25	7.12
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2	0.5661	0.5658	0.6209	0.6202	0.6890	0.6885
Sample size	1711	1711	1711	1711	1711	1711
Predicted impact of increasing class size from:						
10 to 30	-0.16		-0.21		-0.21	
30 to 50	-0.14		-0.15		-0.16	
5 to 15	-0.09	-0.13	-0.12	-0.16	-0.12	-0.07
15 to 25	-0.08	-0.03	-0.11	-0.08	-0.10	-0.11
25 to 35	-0.07	-0.07	-0.09	-0.06	-0.09	-0.06
35 to 45	-0.07	-0.07	-0.08	-0.08	-0.08	-0.09
45 to 55	-0.06	-0.07	-0.06	-0.07	-0.07	-0.05
55 to 65	-0.06	-0.06	-0.05	-0.06	-0.05	-0.09

Parentheses indicate heteroskedastic-consistent standard errors; bold coefficients are statistically significant at the 5% level. All models are weighted by the square root of the number of responses per course, and include indicators for academic department, semester, and year. Columns 1 and 2 include instructor controls, Columns 3 and 4 include both instructor and course controls, while Columns 5 and 6 include instructor, course, and instructor-course controls.

Table 4: Impact of enrollment size on average instructor ratings, R_{tci}

	R_{tci} =rating on statement S1: "Thought provoking ideas and concepts were introduced."			R_{tci} =rating on statement S2: "The instructor encouraged student involvement/questions."		
	(1)	(2)	(3)	(4)	(5)	(6)
Class size	**-0.0121 (0.0039)**			**-0.0146 (0.0037)**		
Class size $^2/100$	0.0059 (0.0045)			**0.0093 (0.0043)**		
Class size 20-39		**-0.1240 (0.0455)**			**-0.1948 (0.0432)**	
Class size 40-59		**-0.2279 (0.0518)**			**-0.2735 (0.0493)**	
Class size 60+		**-0.3558 (0.0733)**			**-0.3742 (0.0697)**	
Class size 10-19			-0.0852 (0.0971)			-0.0189 (0.0925)
Class size 20-29			-0.1505 (0.0972)			**-0.1838 (0.0926)**
Class size 30-39			**-0.2258 (0.0964)**			**-0.2277 (0.0918)**
Class size 40-49			**-0.3058 (0.0989)**			**-0.2883 (0.0942)**
Class size 50-59			**-0.3691 (0.1085)**			**-0.3550 (0.1033)**
Class size 60-69			**-0.4532 (0.1201)**			**-0.3953 (0.1143)**
Class size 70+			**-0.4487 (0.1244)**			**-0.4186 (0.1184)**
F-statistic	5.88	5.81	5.78	6.82	6.76	6.70
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2	0.6424	0.6404	0.6422	0.6756	0.6744	0.6753
Sample size	1711	1711	1711	1711	1711	1711
Predicted impact of increasing class size from:						
10 to 30	-0.19	-0.12		-0.22	-0.19	
30 to 50	-0.15	-0.10		-0.14	-0.08	
5 to 15	-0.11		-0.09	-0.13		-0.02
15 to 25	-0.10		-0.07	-0.11		-0.16
25 to 35	-0.09		-0.08	-0.09		-0.04
35 to 45	-0.07		-0.08	-0.07		-0.06
45 to 55	-0.06		-0.06	-0.05		-0.07
55 to 65	-0.05		-0.08	-0.03		-0.04

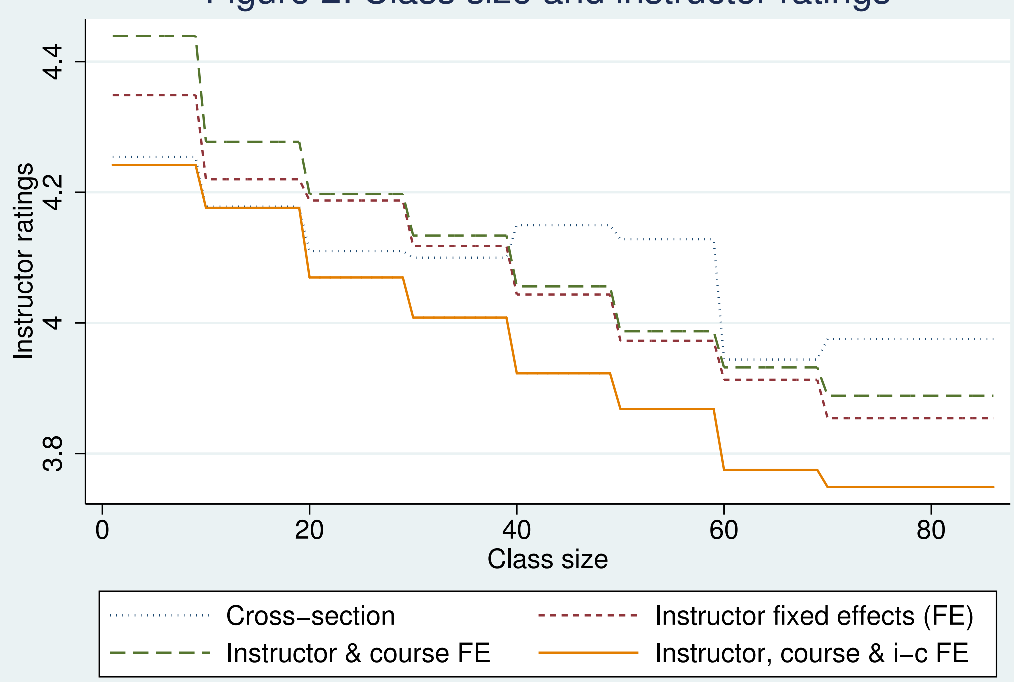
Table 4, continued: Impact of enrollment size on average instructor ratings, R_{tci}

	R_{tci} =rating on statement S3: "The instructor challenged me to think critically."			R_{tci} =rating on statement S4: "Overall the course was effective in helping me learn."		
	(1)	(2)	(3)	(4)	(5)	(6)
Class size	**-0.0102 (0.0038)**			**-0.0141 (0.0045)**		
Class size ^{2}/100	0.0032 (0.0044)			0.0057 (0.0052)		
Class size 20-39		**-0.1404 (0.0451)**			**-0.1565 (0.0529)**	
Class size 40-59		**-0.2377 (0.0513)**			**-0.2965 (0.0602)**	
Class size 60+		**-0.3923 (0.0727)**			**-0.4932 (0.0853)**	
Class size 10-19			-0.1109 (0.0963)			-0.0478 (0.1130)
Class size 20-29			**-0.2081 (0.0964)**			-0.1468 (0.1131)
Class size 30-39			**-0.2528 (0.0956)**			**-0.2285 (0.1122)**
Class size 40-49			**-0.3348 (0.0981)**			**-0.3472 (0.1151)**
Class size 50-59			**-0.3834 (0.1076)**			**-0.3866 (0.1262)**
Class size 60-69			**-0.4751 (0.1191)**			**-0.5438 (0.1397)**
Class size 70+			**-0.5352 (0.1233)**			**-0.5701 (0.1447)**
F-statistic	5.88	5.82	5.77	7.48	7.43	7.37
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2	0.6422	0.6406	0.6417	0.6956	0.6946	0.6957
Sample size	1711	1711	1711	1711	1711	1711
Predicted impact of increasing class size from:						
10 to 30	-0.18	-0.14		-0.24	-0.16	
30 to 50	-0.15	-0.10		-0.19	-0.14	
5 to 15	-0.10		-0.11	-0.13		-0.05
15 to 25	-0.09		-0.10	-0.12		-0.10
25 to 35	-0.08		-0.04	-0.11		-0.08
35 to 45	-0.08		-0.08	-0.10		-0.12
45 to 55	-0.07		-0.05	-0.08		-0.04
55 to 65	-0.06		-0.09	-0.07		-0.16

Parentheses indicate heteroskedastic-consistent standard errors; bold coefficients are statistically significant at the 5% level. All models are weighted by the square root of the number of responses per course, and include indicators for academic department, semester, and year, as well as instructor, course, and instructor-course fixed effects.

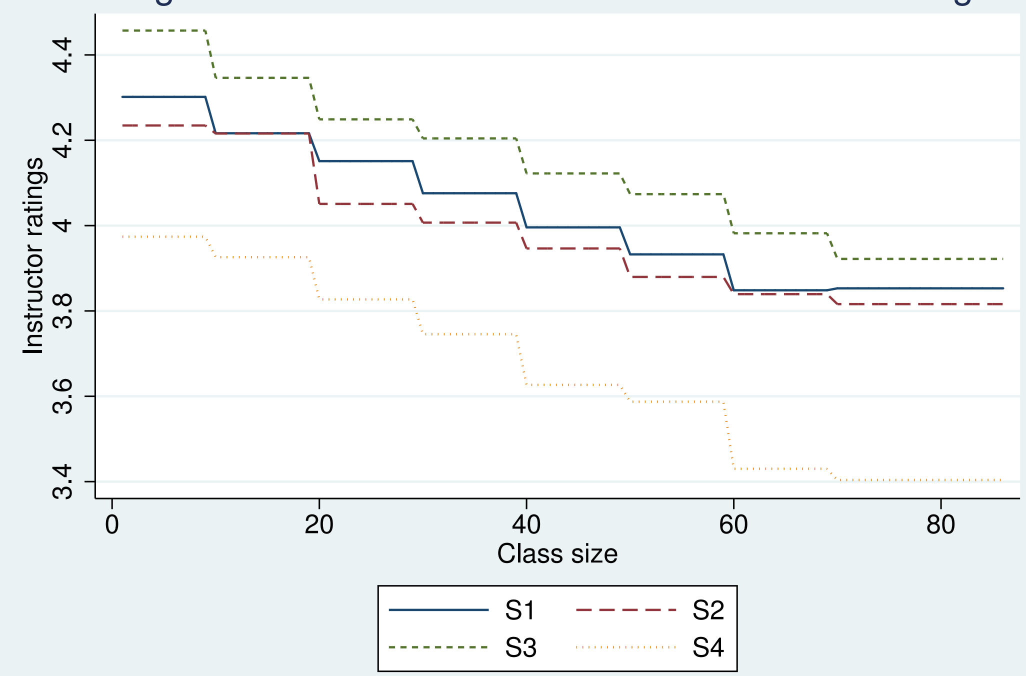
To facilitate understanding of several key aspects highlighted by the differences between the specifications described above, I compare the four specifications graphically in Figure 2. Curves in Figure 2 are drawn for a course in the largest of four departments in the 2014 spring semester. First, the apparent difference between the cross-section and instructor fixed effects results likely captures the potential that department heads may assign better instructors to larger sections. This difference is easily seen with class sizes from 40 to 60, as well as in class sizes above 70. Second, when adding course fixed effects to the instructor fixed effects, results for the point estimates are slightly more positive. This exaggeration of the estimates in the other specifications is likely due to the fact that often more difficult classes will attract students who are more motivated, yet also be associated with higher instructor ratings and smaller class sizes. However, without the additional control for course differences, this exaggeration cannot be accounted for. Finally, the results from the ideal specification including instructor, course, and instructor-course fixed effects show that even beyond controlling for instructor and course effects, it is useful to control for instructor-course interaction as a potential important driver of outcomes (for example, an instructor could be particularly skilled at teaching a subset of the assigned courses).

Figure 2: Class size and instructor ratings



Given that the dependent variable for all estimations of Equations 1 and 2 is a combination of the results from four selected ratings on teaching evaluations, it is worth checking if this selection unduly influences the results. As checks for robustness, I reestimate Equations 1 and 2 using R_{tci} as a placeholder for each of the four separate evaluation statement ratings in place of Z_{tci}. Results for $S1$, $S2$, $S3$, and $S4$ are presented in Table 4, highlighting the preferred specifications using instructor, course, and instructor-course fixed effects. While results clearly vary given the different dependent variables in placeholder R_{tci}, the relationships outlined by the results in Table 3 are generally confirmed by those in Table 4. Figure 3 shows the predicted evaluation ratings for a given enrollment (again, using a course in the largest of the four departments in the 2014 spring semester), comparing all four measures in the preferred specification where indicators group class sizes by tens. While the associated drop in ratings with an increase in class size clearly varies slightly across the different rating measures, the prevailing negative relationship between enrollment and instructor ratings is consistently strong across all four measures. All four measures have a decrease that is not statistically significant when increasing class size from 0-9 to 10-19, however all the remaining increases across the enrollment grouped by tens are associated with significant decreases in instructor ratings (with the exception of S1 and S4, moving from 10-19 to 20-29).

Figure 3: Class size and alternative instructor ratings



As an additional robustness check, I follow Karas (2019) in performing all estimations in Stata with the routine *reghdfe* (Correia, 2017). This addresses the potential of issues arising from the combination of fixed effects and the inclusion of singletons (Correia, 2015). As seen in the results highlighted in Table C in the Appendix, this exclusion of singletons results in only minor variation regarding the magnitude and significance of estimates compared to the baseline results from Table 3.

5. Conclusion

Examining the connection between class enrollment size and instructor evaluation ratings, I find a consistently negative relationship that is statistically significant across numerous specifications that vary both in functional form and inclusion of fixed effects. Focusing on the results from the preferred specification, where class size indicators group enrollment by tens and fixed effects control for differences in instructors, courses, and even instructor-course combinations, increasing class size from 0-9 to 10-19 is associated with an average decrease of 0.07 for instructor ratings. This decrease increases slightly to 0.11, followed by decreases of 0.06, 0.09, 0.05, and 0.09 points, respectively, as class size is increased over the next six subsequent enrollment groups of ten, with the final decrease of 0.03 points associated with an increase in class size from 60-69 to greater than 70.

Interestingly, the comparison of cross-sectional specifications with specifications including instructor fixed effects, instructor and course fixed effects, and the preferred specification provides support for the notion that (1) department heads may tend to assign better instructors to larger sections and (2) more difficult classes tend to attract students who are more motivated, yet also be associated with higher instructor ratings and smaller class sizes, and (3) instructors may be particularly skilled at teaching a subset of assigned courses. The overall estimated negative relationship sheds light on what must be considered an important aspect in further understanding of student engagement, a topic clearly important to students, faculty, and administrators in higher education. Especially given the prevalence of measuring engagement at least partially with instructor evaluation ratings, while external validity cannot be claimed given the focus on my specific sample, the results do point to the necessity of working toward a better understanding of how class size may generally affect evalution ratings through further studies.

References

Beattie, I., & Thiele, M. (2016). Connecting in class? College class size and inequality in academic social capital, The Journal of Higher Education, 87(3), 332-362.

Bedard, K., & Kuhn, P. (2008). Where class size really matters: class size and student ratings of instructor effectiveness, *Economics of Education Review*, 27, 253-265.

Carini, R., Kuh, G., & Klein, S. (2006). Student engagement and student learning: testing the linkages, *Research in Higher Education*, 47(1), 1-32.

Correia, S. (2015) Singletons, cluster-robust standard error and fixed effects: a bad mix, Technical Note, Duke University.

Correia, S. (2017) reghdfe: stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects, Statistical Software Components S457874, Boston College Department of Economics.

DeCanio, S. (1986). Student evaluations of teaching - a multinomial logit approach, *Journal of Economic Education*, 17, 165-175.

Karas, A. (2019). The effect of class size on grades and course evaluations, Working Paper, Utrecht University.

McConnell, C., & Sosin, K. (1984). Some determinants of student attitudes toward large classes, *Journal of Economic Education*, 38, 220-229.

Monks, J., & Schmidt, R. (2011). The impact of class size on outcomes in higher education, The B.E. Journal of Economic Analysis & Policy, 11(1), 1-19.

Ragan, J., & 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.

Sapelli, C., & Illanes, G. (2016). Class size and teacher effects in higher education, Economics of Education Review, 52(C), 19-28.

Shin, Y. (2011). The causal effect of class size on academic achievement, Journal of Educational and Behavioral Statistics, 36(2), 154-185.

Siegfried, J., & Kennedy, P. (1995). Does pedagogy vary with class size in introductory economics? *American Economic Review, Papers and Proceedings*, 85, 347-351.

Appendix

Table A: Impact of enrollment size on average instructor ratings, Z_{tci}

	Cross-section (no fixed effects)		
	(1)	(2)	(3)
Class size	-0.0045 (0.0039)		
Class size ^{2}/100	0.0028 (0.0048)		
Class size 20-39		**-0.0920 (0.0484)**	
Class size 40-59		-0.0489 (0.0532)	
Class size 60+		**-0.2345 (0.0776)**	
Class size 10-19			-0.0762 (0.1054)
Class size 20-29			-0.1443 (0.0998)
Class size 30-39			-0.1543 (0.0960)
Class size 40-49			-0.1045 (0.0984)
Class size 50-59			-0.1260 (0.1110)
Class size 60-69			**-0.3102 (0.1276)**
Class size 70+			**-0.2787 (0.1301)**
F-statistic	8.75	8.60	6.19
P-value	0.0000	0.0000	0.0000
R^2	0.0442	0.0482	0.0486
Sample size	1711	1711	1711
Predicted impact of increasing class size from:			
10 to 30	-0.07	-0.09	
30 to 50	-0.05	0.04	
5 to 15	-0.04		-0.08
15 to 25	-0.03		-0.06
25 to 35	-0.03		-0.01
35 to 45	-0.03		0.05
45 to 55	-0.02		-0.03
55 to 65	-0.01		-0.18

Parentheses indicate heteroskedastic-consistent standard errors; bold coefficients are statistically significant at the 5% level. All models are weighted by the square root of the number of responses per course, and include indicators for academic department, semester, and year.

Table B: Impact of enrollment size on average instructor ratings, Z_{tci}			
	Instructor fixed effects	Instructor and course fixed effects	Instructor, course, and instructor-course fixed effects
	(1)	(2)	(3)
Class size 20-39	**-0.1081 (0.0373)**	**-0.1397 (0.0426)**	**-0.1539 (0.0428)**
Class size 40-59	**-0.2117 (0.0429)**	**-0.2419 (0.0490)**	**-0.2589 (0.0488)**
Class size 60+	**-0.3531 (0.0638)**	**-0.3668 (0.0702)**	**-0.4039 (0.0690)**
F-statistic	12.39	8.05	7.17
P-value	0.0000	0.0000	0.0000
R^2	0.5630	0.6179	0.6872
Sample size	1711	1711	1711
Predicted impact of increasing class size from:			
10 to 30	-0.11	-0.14	-0.15
30 to 50	-0.10	-0.10	-0.10

Parentheses indicate heteroskedastic-consistent standard errors; bold coefficients are statistically significant at the 5% level. All models are weighted by the square root of the number of responses per course, and include indicators for academic department, semester, and year. Column 1 includes instructor controls, Column 2 includes both instructor and course controls, while Column 3 includes instructor, course, and instructor-course controls.

Table C: Impact of enrollment size on average instructor ratings, Z_{tci}(no singletons)						
	Instructor fixed effects		Instructor and course fixed effects		Instructor, course, and instructor-course fixed effects	
	(1)	(2)	(3)	(4)	(5)	(6)
Class size	**-0.0010 (0.0044)**		**-0.0148 (0.0045)**		**-0.0120 (0.0051)**	
Class size ^{2}/100	0.0031 (0.0052)		0.0084 (0.0045)		0.0053 (0.0051)	
Class size 10-19		-0.1488 (0.1036)		-0.1666 (0.0857)		-0.0505 (0.1253)
Class size 20-29		-0.1768 (0.1093)		**-0.2515 (0.0980)**		-0.1557 (0.1291)
Class size 30-39		**-0.2446 (0.1065)**		**-0.3170 (0.1041)**		-0.2223 (0.1373)
Class size 40-49		**-0.3331 (0.1114)**		**-0.4083 (0.1153)**		**-0.3090 (0.1449)**
Class size 50-59		**-0.3980 (0.1215)**		**-0.4841 (0.1175)**		**-0.3594 (0.1533)**
Class size 60-69		**-0.4436 (0.1299)**		**-0.5239 (0.1519)**		**-0.4380 (0.1738)**
Class size 70+		**-0.5240 (0.1182)**		**-0.5791 (0.1274)**		**-0.4772 (0.1635)**
F-statistic	16.09	5.33	11.06	6.50	10.08	4.92
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
R^2	0.5080	0.5077	0.5917	0.5879	0.6313	0.6308
Sample size	1693	1693	1670	1670	1582	1582
Predicted impact of increasing class size from:						
10 to 30	-0.17		-0.20		-0.20	
30 to 50	-0.14		-0.15		-0.15	
5 to 15	-0.09	-0.15	-0.11	-0.17	-0.11	-0.05
15 to 25	-0.08	-0.03	-0.10	-0.08	-0.10	-0.11
25 to 35	-0.08	-0.07	-0.09	-0.07	-0.09	-0.07
35 to 45	-0.07	-0.09	-0.08	-0.09	-0.08	-0.09
45 to 55	-0.06	-0.06	-0.07	-0.08	-0.07	-0.05
55 to 65	-0.06	-0.05	-0.06	-0.04	-0.06	-0.08

Parentheses indicate heteroskedastic-consistent standard errors; bold coefficients are statistically significant at the 5% level. All models are weighted by the square root of the number of responses per course, and include indicators for academic department, semester, and year. Columns 1 and 2 include instructor controls, Columns 3 and 4 include both instructor and course controls, while Columns 5 and 6 include instructor, course, and instructor-course controls.