3-2011

The Impact of Class Size on Outcomes in Higher Education

James Monks
University of Richmond, jmonks@richmond.edu

Robert M. Schmidt
University of Richmond, rschmidt@richmond.edu

Recommended Citation
Monks, James, and Robert M. Schmidt. "The Impact of Class Size on Outcomes in Higher Education." *The B.E. Journal of Economic Analysis & Policy* 11, no. 1 (March 2011): Article 62. doi:10.2202/1935-1682.2803.

The B.E. Journal of Economic Analysis & Policy

Contributions

Volume 11, *Issue* 1 2011 *Article* 62

The Impact of Class Size on Outcomes in Higher Education

James Monks* Robert M. Schmidt†

*University of Richmond, jmonks@richmond.edu
†University of Richmond, rschmidt@richmond.edu

Recommended Citation
James Monks and Robert M. Schmidt (2011) "The Impact of Class Size on Outcomes in Higher Education," *The B.E. Journal of Economic Analysis & Policy*: Vol. 11: Iss. 1 (Contributions), Article 62.
Available at: http://www.bepress.com/bejeap/vol11/iss1/art62

The Impact of Class Size on Outcomes in Higher Education

James Monks and Robert M. Schmidt

Abstract

Numerous studies have investigated the impact of class size on student outcomes. This analysis contributes to this discussion by isolating the impact of class size on student outcomes in higher education by utilizing a natural experiment at a selective institution which enables the estimation of class size effects conditional on the total number of students taught by a faculty member. We find that class size negatively impacts student assessments of courses and instructors. Large classes appear to prompt faculty to alter their courses in ways deleterious to students.

KEYWORDS: class size, student outcomes, education production function

I. Introduction

Numerous studies have investigated the influence of class size on student attitudes, behaviors, and outcomes. The overwhelming majority of these studies have focused on elementary school and even pre-school effects of class size on student achievement. The conventional wisdom among parents, teachers, school administrators, and policy makers is that smaller class sizes translate to improvements in student learning and outcomes. This conventional wisdom, however, has not been universally supported by empirical evidence. While a number of studies have found support for the importance of class size on student achievement, others strongly refute this claim concluding that class size has little to no impact on objective student outcomes. The difficulties in assessing the causal influence of class size on student outcomes, such as achievement, are (1) class size itself is often not directly observed but rather proxied by pupil-teacher ratios at the state, district, or school level, (2) many data sets used to analyze this question are cross-sectional and thus do not allow one to control for fixed student, teacher, class, or school effects, and (3) class size itself may be endogenous in a student outcome equation. Nonetheless, the general consensus among researchers examining this issue is that if class size matters at all its influence is most pronounced at the lowest grade levels.

Only a handful of studies have focused on the role that class size may play in outcomes in tertiary education. Clearly, the educational environment is dramatically different from the classroom and learning environment of the elementary school setting. Even so the conventional wisdom of the benefits of small class size persists in postsecondary education, as well. This intractable perception is so prevalent that class size represents two of the fifteen inputs into the *U.S. News and World Report* college rankings formula, despite the lack of convincing evidence that class size has a significant impact on student outcomes. A university’s rank is a function of the percentage of course sections that it offers with fewer than twenty students and the percentage of course sections that it offers with fifty or more students. The former enters the rankings formula positively, and the latter negatively. While the *U.S. News and World Report*, and other college rankings, are often criticized for their focus on inputs rather than outputs in the educational process, rarely is the use of class size the primary focus of this criticism.

Not only is the educational setting dramatically different in tertiary education in comparison to elementary education, the primary focus of most of this literature, but class size itself may have a fundamentally different relationship to student outcomes in postsecondary education. In elementary school, where a teacher usually teaches the same class of students almost all school day long, class size and the number of students that a teacher is responsible for are equivalent. In higher education, a professor may teach one, two, three or more sections of a course each semester. In this case, class size -- the number of students in a class section -- and the total number of students that a professor is responsible for may or may not be the same; thus the existing studies of class size in higher education combine and confound class size effects and total student responsibility effects.

This study overcomes many of the shortcomings outlined above by taking advantage of a unique policy change within a business school at a private, selective university in the United States. A new dean of business at this university allowed professors who were teaching three sections of a course per semester to "super-size" these sections into two larger sections, of roughly equal total number of students. For example, a typical super-sized course went from three sections with a cap on enrollment per section of thirty students to two sections capped at forty-five students. The impetus for this policy was an attempt to lower the teaching responsibilities of tenure-track faculty without increasing the use of adjunct faculty. This practice persisted for approximately six years, until a new dean arrived at the school. With an eye on the new *Business Week* undergraduate business school rankings, which are a function of average class size, the dean eliminated the use of super-sizing and returned those faculty to three sections. This natural experiment allows us to compare student outcomes before, during, and after significant changes in class size as a result of the policy change and subsequent reversal, all the while maintaining the same policy regarding the overall student load of each faculty member. This stands in contrast to policy changes made for cost-cutting reasons where class sizes are raised without proportionally reducing the number of sections taught. Since both class size and student load rise, their effects cannot be separated. By contrast, any impacts discernible during this policy are attributable entirely to changes in section size.

Additionally, because we are able to track the same professors over time, we are able to control for faculty and course fixed effects and thus estimate within instructor and course class size effects. As not all faculty were eligible for this policy change, because they taught fewer than three sections of the same course per semester, and because not all faculty opted to super-size even if eligible, we are able to compare the student outcomes of the super-size class sections to a control group of faculty over the same time period in the same courses. Furthermore, because the super-size policy did not significantly alter the total number of students that a faculty member was responsible for teaching, this policy allows us to identify the direct effect of class size on student outcomes separate from faculty student-load effects.

This paper will examine the influence of significant changes in class size on student assessments of university business courses conditional on fixed instructor and course effects benchmarked against instructors and courses at the same university whose class sizes were not substantially altered over the same time period. The following section of the paper will briefly outline the literature on the impact of class size on student achievement in higher education. This section is followed by a discussion of the data and then direct empirical tests of the relationship between class size and student outcomes. Finally, the conclusion of the paper discusses the implications of these results.

II. Literature Review

As mentioned above there is a vast literature on the role of class size in student achievement.[1] The bulk of this literature focuses on whether class size is instrumental in improving learning and academic achievement at the elementary school level. Analyses have been based on randomized experiments (see, for example Krueger, 1999; and Achilles et al., 1995), discontinuities in class size (e.g., Angrist and Lavy, 1997; and Hoxby, 2000), as well as the National Educational Longitudinal Study (NELS) (e.g., Goldhaber and Brewer, 1997; and Akerhielm, 1995). The results are mixed. Many studies find a significant negative impact of larger class sizes, others find no significant impact, and surprisingly some find a significant positive impact once endogeneity is accounted for in the NELS data.

The studies most relevant to this paper examine the influence of class size on student outcomes in tertiary education. Two studies provide a theoretical basis for the role of class size and overall student load on student outcomes. Correa (1993) focuses on the role of individual faculty-student interaction. His model describes teachers that must weigh efforts directed to students as a group rather than to individual students. The larger the class and/or overall student load the greater is the instructor effort devoted to class-wide activities at the expense of individual attention. In this way, individual student learning and outcomes decline as class size and/or student load increases. His model illustrates the importance of separating class size effects from total student load effects. Lazear (2001) outlines a theoretical model where class size itself is important due to the role that class size plays in setting the classroom environment. Large classes (more students) may allow students to be more disruptive, allow them to "hide" from participation, engagement, or even attendance, while small classes may more easily lend themselves to pedagogical activities that improve learning, such as hands on activities and student-faculty classroom interaction.

Empirical studies of the role of class size in higher education face two challenges: (1) the lack of standardized tests across multiple instructors in the

[1] For a more complete discussion of the influence of class size on primary and secondary student achievement, see Averett and McLennan (2004).

same course and (2) the lack of randomness in student assignment to the different sections. Because most instructors in higher education cannot be made or incentivized to administer a single standardized test, the most commonly examined measures are student grades in a course and student assessment of the quality of a course or instructor. Studies that have examined student assessment of a course universally agree that class size has a negative impact on student course evaluations, with larger courses receiving statistically significant lower scores than smaller courses. For example, Bedard and Kuhn (2008) examine student evaluations of economics courses at the University of California at Santa Barbara, from 1997 to 2004. They find a large, highly significant, and nonlinear negative impact of class size on student evaluations of instructor effectiveness. Their result is robust to instructor and course fixed effects. Similarly, Walia (2008) utilizes nineteen semesters of student evaluations of economics courses at Kansas State University. Once again class size is found to have a negative and statistically significant impact on student course evaluations. Illustrating that this result is not unique to economics students nor the United States, Westerlund (2008) reports that increases in the size of mathematics classes at Lund University in Sweden lead to significantly lower student course evaluations there, as well.

Fewer studies still have examined the impact of class size on student performance in higher education. Bandiera et al. (2010) examine administrative records from a leading UK university and find a significant negative, but highly non-linear effect of class size on student test results. They conclude that changes in class size have a significant impact on student performance but only at the very top and bottom of the class size distribution. Furthermore, they find that students at the top of the grade distribution are most negatively affected by class size, particularly in large class sections. They rule out class size effects being due to non-random assignment of faculty across class size, student self-selection into class size, omitted inputs, or changes in grading policies based on class size. Similarly, Kokkelenberg et al. (2008) find that average grades decline significantly with class size at a public northeastern U.S. university. They find that grades drop dramatically with class size up to twenty students, and less steeply but nonetheless monotonically thereafter.

While the existing literature on primary and secondary class size effects is mixed, the evidence of class size on student assessment and student grades in higher education is more consistent. Unfortunately, the higher education literature suffers from a lack of random, substantial changes in class size within instructors and courses, and confounds class size effects with student oversight effects relating to total student load. This paper contributes to this literature by exploiting a natural experiment in class sizes that allows for an approximate fifty percent increase in class size, within instructor and course, without a change in the total number of students taught by that instructor.

DOI: 10.2202/1935-1682.2803

III. Data and Econometric Model

The data for this study come from administrative records and student course evaluations at a private, highly selective university on the east coast of the United States. The faculty and courses utilized in this analysis are restricted to the undergraduate business school within this university, as the student course evaluation instrument is specific to the business school and the practice of super-sizing outlined above did not include graduate courses. The sample period covers the academic years 1996 through 2008. This allows for three years of data preceding the implementation of the super-size policy, six years of super-sizing, and three years following the suspension of this practice. The sample includes 48 individual faculty members, 88 separate courses, and 1,928 course sections. In total, 12 faculty, 8 courses, and 80 sections were super-sized over this period.

The outcomes available for and examined in this paper are the average course ratings from the student course evaluations for: (1) overall instructor rating (1-5, with 5 being the best); (2) amount learned (1-5, with 5 being the most); (3) overall course rating (1-5, with 5 being the best), and (4) expected course grade (calculated as the average of two questions asking students their lowest and highest expected grade in the course). The results in this paper must be qualified accordingly – differences in student perceptions might or might not translate into differences in academic outcomes. Nevertheless, student course evaluations are studied widely for at least three reasons. First, objective student outcomes, such as standardized test results, rarely are available to address issues such as the impact of varying class size. Second, studies such as Kulik (1999) find that student ratings generally correlate highly with other accepted measures of academic outcomes. Third, from the perspective of the faculty, subjective student course evaluations are perhaps even more important as this is the primary information used in assessing and rewarding instructors for their teaching at this and many institutions.

The explanatory variables fall into several categories. First, course specific variables such as the average grade point average of the students in the class, the gender mix of the class (proportion male and an interaction of proportion male with a dummy variable indicating the professor was male), the grade level mix of the class (proportion seniors, juniors, sophomores, and freshmen), the meeting time of the class (early or late), class length (50 minutes three times per week versus 75 minutes twice a week), and the prior level of student interest in the course. Additionally, two other sets of controls are included for institutional factors: (a) dummy variables capturing time effects (semesters) to allow for observed evaluation inflation and (b) an “online” binary indicating the evaluation was administered online (discussed later).

The second category of explanatory variables captures the number of students. Class size is taken from administrative records that list the actual number of students enrolled in the class after the drop/add period in the second week of a fifteen week semester. We employ a super-size dummy variable (with one indicating that a class section was super-sized) as an instrument for class size to capture the exogenous increase in class size prompted by the change in policy at this institution. A simple difference-in-means test indicates that super-sized sections were 14.8 students larger on average than non-super-sized sections. This translates into an increase in class size of approximately 64.8 percent and is statistically different from zero at the 99 percent level. Controlling for all exogenous variables in the model, the first-stage results for the endogenous variable, class size, indicate that super-sized sections for comparable courses were 13.4 students larger than non-super-sized sections, statistically significant at the 99 percent level (see Table 2 below). Additionally, we are also able to control for student load, the total number of students that each instructor had enrolled in all of his or her class sections that semester. Although student load suffers from the same issues with respect to endogeneity that class size does, we are unable to identify a satisfactory instrument for student load. Correspondingly, although we include student load to control for its separate effect, we do not put much emphasis on interpreting its coefficient.

The final set of explanatory variables includes time, instructor, and course fixed effects. These allow us to control for effects that otherwise would confound the estimation of impacts on average course ratings for those super-sizing their course sections, and thus dramatically increasing their class sizes, to those who did not see significant changes in their class sizes over the same period.

The econometric specification we estimate is:

(1) $$Y_{isjt} = \beta_0 + \beta_1(X_{isjt}) + \beta_2(Class\ Size_{isjt}) + \beta_3(Student\ Load_{it}) + \delta_i + \alpha_j + v_t + \varepsilon_{isjt};$$

where Y represents the average course outcomes outlined above for instructor i, in section s of course j, at time (semester) t. The vector X indicates course specific variables such as average student grade point average which may influence student evaluations, class size indicates the number of students enrolled in the class, and student load is the total number of students taught by the faculty member that semester over all courses. Table 1 presents descriptive statistics for all of these variables.

DOI: 10.2202/1935-1682.2803

Table 1
Summary Measures

	Minimum	Maximum	Mean	Std. Dev.
Class Size	2	45	23.39	6.86
Student Load	10	150	63.46	18.75
Early	0	1	.098	.297
Late	0	1	.284	.451
50 Minute Class	0	1	.312	.463
Proportion Male Students	0	1	.601	.151
Male Student * Male Faculty	0	1	.516	.256
Proportion Sophomores	0	1	.226	.315
Proportion Junior	0	1	.342	.325
Proportion Senior	0	1	.372	.376
GPA	2.75	3.69	3.18	.142
Expected Grade	2.04	3.97	3.05	.292
Interest in This Subject Prior to the Course	1.75	5.00	3.47	.545
Course Workload	2.27	4.92	3.39	.463
Course Level of Difficulty	2.45	5.00	3.66	.501
Level of Critical & Analytical Thinking	2.20	5.00	4.00	.447
Pace of Course	2.20	4.45	3.15	.207
Number of Evals. Completed by Student	1.00	4.67	2.66	.824
Clear and Understandable Presentation	2.00	5.00	4.09	.485
Effectiveness of Teaching Methods	2.04	5.00	4.05	.500
Instructor's Daily Preparation for Class	2.54	5.00	4.50	.361
Effectiveness in Stimulating Interest	2.03	5.00	3.94	.514
Enthusiasm for Teaching Course	2.87	5.00	4.48	.366
Availability Outside the Classroom	2.36	5.00	4.41	.370
Respect for Students in the Class	2.50	5.00	4.34	.425
Graded Material vs. Course Content	1.44	5.00	3.98	.499
Adequacy of Comments on Student Work	1.78	5.00	3.76	.483
Timeliness of Feedback on Student Work	1.26	5.00	4.26	.496
Usefulness of Text, etc. to Learning	1.66	5.00	3.84	.512
Amount Learned as a Result of Course	2.45	5.00	4.12	.449
Instructor's Overall Teaching Ability	2.08	5.00	4.18	.479
Overall Quality of Course	2.30	5.00	4.00	.475
Number of Faculty	48			
Number of Courses	88			
Number of Sections	1,928			

The error components δ, α, ν represent instructor, course, and time effects, respectively. Equation (1) is estimated using weighted two-stage least squares (2SLS) with the super-sized binary used as an instrument for class size and the weights determined in an iterative manner as suggested in Dickens (1990). Specifically, a test for a group error component (in our case, an error component relating to unobservables in each course section) rejected the null hypothesis of no group error component, in favor of the use of weighted least squares to allow for its presence. Accordingly, we regressed the squared residuals from the 2SLS regression on one over N_{isjt}, the number of respondents in the section. We weighted each observation by one over the square root of the predicted value from that regression. These two regressions were estimated iteratively until coefficient estimates from the regression of squared residuals converged up to three decimal places.

Before proceeding, it is important to consider who among the faculty chose to super-size (versus those who did not). If in fact super-sizing, or dramatically increasing class size, does have a deleterious effect on course outcomes, then it is safe to assume that faculty who felt that they would be least negatively affected by the increase in class size would be more likely to opt to super-size their course sections. Similarly, faculty who felt that their course ratings would be most negatively affected by super-sizing would be less likely to pursue this option. If this is the case, our estimates of the effects of enrollment on course outcomes provide a lower bound of the impact of class size on course evaluations. Alternatively, in the face of uncertainty of its effects, it may be the case that faculty for whom a decline in course evaluations would have the least consequence would be more likely to super-size. For example, tenure track (but not yet tenured) faculty would be the most at risk from a decline in course evaluations, and would likely opt not to utilize this option. If this is the case, then our estimates of the impact of class size on course outcomes are likely accurate. In fact, of the eleven faculty who super-sized their course sections all but one were tenured faculty at the time. The above arguments imply that our estimates based on this natural experiment, if anything, provide a lower bound of the impact of class size on course outcomes.

IV. Empirical Results

We begin by examining the influence of class size on the self-reported outcomes of: (1) how much the students reported learning in the course; (2) overall instructor rating for the course; (3) overall course rating; and (4) the average expected grade for the course. These four measures are used to gauge the impact of class size on course outcomes as assessed by the students. As noted above,

DOI: 10.2202/1935-1682.2803

although student ratings are used widely in assessing faculty, student ratings may or may not translate into actual learning differences.

Table 2 presents the results of the first-stage regression on class size as well as the weighted 2SLS regressions for the four outcomes described above.[2] These regressions control for dummy variables for whether it was an early morning class (classes that begin before 9AM), late afternoon class (classes that begin after 2PM), met three times a week for 50 minutes (versus twice a week for 75 minutes), and whether the evaluations were administered online versus in paper (faculty had the option of using online evaluations the last two semesters of the sample period). Additional controls include the proportion of the class that was male, an interaction of male professor and the proportion of male students, the proportion in each grade (sophomore, junior, senior), the average self-reported GPA of the class, and the reported level of interest in the subject matter prior to the course. We also control for the student load (total number of students taught by that instructor that semester), so that the coefficient on class size solely reflects the impact of class size on course outcomes and not total student responsibility. All regression results include time, course, and faculty fixed effects; so that all estimated coefficients represent within course and faculty effects of changes in course enrollment on student outcomes.

The first-stage regression's R^2 of 66% is driven by the statistical significance of over 80 fixed effects binaries and three additional significant variables. The super-sized binary appears to be an effective instrument with a t-value exceeding twenty-three, 99% significance, and a coefficient indicating that super-sized sections averaged 13.4 more students than comparable sections. Students tend to shy away from early classes (significant at 99%). Somewhat surprisingly, especially given the inclusion of course and faculty binaries, higher GPA students tend toward larger sections of the same course.

[2] Although these are the first-stage results for the amount learned regression, the results are very similar across the four output measures, but do vary due to the weighting of both the first and second stage regressions. For example, t-values for the super-size binary are over 20 throughout and its coefficient is 13.36, 13.38, 13.36, and 13.34, respectively.

Table 2
Regression Results
Impact of Class Size on Course Outcomes

	1st-Stage Regression	How Much Learned	Instructor Rating	Course Rating	Expected Grade
Intercept	16.810 ***	3.161 ***	3.584 ***	2.878 ***	0.553 ***
Super-sized Dummy	13.432 *** (0.583)				
Class Size		-0.0083 *** (0.0029)	-0.0089 *** (0.0034)	-0.0095 *** (0.0032)	-0.0037 ** (0.0015)
Student Load		-0.0021 *** (0.0005)	-0.0019 *** (0.0006)	-0.0018 *** (0.0005)	0.0005 * (0.0003)
Early	-2.873 *** (0.380)	-0.072 *** (0.026)	-0.088 *** (0.030)	-0.094 *** (0.029)	-0.058 *** (0.015)
Late	-0.401 (0.273)	0.006 (0.018)	0.004 (0.021)	0.015 (0.020)	0.012 (0.010)
50 Minute Class	-0.362 (0.303)	0.042 ** (0.020)	0.037 (0.023)	0.044 ** (0.022)	0.015 (0.011)
Online Evaluation	1.201 (1.271)	0.018 (0.083)	-0.026 (0.096)	-0.063 (0.092)	-0.095 * (0.049)
Proportion: Male Students	-0.770 (1.338)	-0.056 (0.087)	-0.224 ** (0.101)	-0.230 ** (0.096)	0.044 (0.051)
Male Student * Male Faculty	-0.186 (1.316)	0.027 (0.085)	0.224 ** (0.099)	0.192 ** (0.095)	0.020 (0.050)
Proportion: Sophomore	0.259 (0.948)	0.161 *** (0.062)	0.218 *** (0.073)	0.253 *** (0.068)	0.150 *** (0.035)
Junior	-0.170 (1.236)	0.348 *** (0.080)	0.472 *** (0.094)	0.490 *** (0.089)	0.204 *** (0.046)
Senior	1.818 (1.418)	0.306 *** (0.092)	0.554 *** (0.108)	0.514 *** (0.102)	0.234 *** (0.053)
GPA	2.564 *** (0.986)	-0.150 ** (0.064)	-0.179 ** (0.075)	-0.191 *** (0.071)	0.573 *** (0.038)
Prior Interest	-0.197 (0.442)	0.413 *** (0.029)	0.352 *** (0.033)	0.461 *** (0.032)	0.123 *** (0.017)
R^2	0.655	0.665	0.597	0.633	0.738
Adjusted R^2	0.621	0.632	0.558	0.598	0.713

Notes: Faculty, semester and course fixed effects were included in the regressions but not in this table. Standard errors are shown in parentheses; * indicates two-tailed significance at the 90% level, ** at 95%, and *** at 99%.

DOI: 10.2202/1935-1682.2803

Turning to the results for course outcomes, consider first the control variables that are significant at the 99% level in all four equations. Faculty interested in maximizing their course evaluation ratings should (a) avoid early morning sections as these consistently earn lower ratings in the four outcomes; (b) attempt to attract upper-class students, especially juniors and seniors; and (c) attempt to attract students who have a high level of prior interest in the course. Of course, students with a high level of a priori enthusiasm for a course might be expected to work harder, learn more and have a better appreciation for the instructor and course. On the other hand, the statistical and quantitative strength of the class binaries and the interest variable is particularly noteworthy given the inclusion of controls for course and faculty fixed effects. These control for required versus elective courses, pre-major versus courses in the major, as well as courses oriented toward freshmen and sophomores versus juniors and seniors.

GPA also has a significant impact on student ratings for all four measures at least at the 95% level. The higher the average GPA of the class the lower the reported amount learned, instructor rating, and course rating even though, as expected, the higher the average GPA the higher the expected grade in the class. The remaining control variables have mixed or no significant impact. Courses that meet three times a week for fifty minutes receive higher average ratings for the amount learned and course (95% level) but not for the instructor nor for expected grade. It is also interesting to note that the higher the proportion of the class that is male the lower the average rating for the instructor and course rating for female professors, but not for male professors − the negative coefficient for proportion male is offset almost entirely by the positive interaction term for proportion male times the male instructor binary.

Given its likely endogeneity, we do not wish to overly interpret the student load variable. Nevertheless, we note that within these regressions, student load is negative and statistically significant at the 99% level for student ratings of the amount learned, instructor, and course. However, at the 90% level, teaching more students in a semester appears to raise students' expected grade.

Finally, turning to the influence of class size on course outcomes we find that the larger the section size, the lower the self-reported amount learned, the instructor rating, and the course rating at the 99% level, and the lower the expected grade at the 95% level. Clearly, students feel that they learn less and get less out of large class sections, even conditional on the number of total students for which an instructor is responsible.

As mentioned above, this institution experimented in the middle years of this sample with combining three sections of a course into two larger sections. This introduced substantial increases in class size, both within faculty members and within courses. This consolidation usually involved increasing class size enrollment caps by about sixty-five percent, or as outlined above by an average of approximately 13 students per super-sized section for comparable courses. This translates to an average decrease in student ratings of about twenty-five percent of a standard deviation for amount learned, instructor quality and course quality.

Additionally, we experimented with interacting the super-size dummy variable with a super-sizing trend (number of semesters having taught super-sized sections) or dummy variables indicating number of semesters having taught super-sized sections to investigate if the significant, negative impacts of class size found above diminished as an instructor became more experienced handling larger class sections. We found no significant evidence of diminution (or amplification) of the negative effects of class size on student outcomes over time.

We also compared our empirical results to those from Bedard and Kuhn (2008), who employ a model similar to that in Table 3. Both studies model instructor effectiveness measured on a five-point scale, focus on the effects of class size, include instructor and course fixed effects, include at least one additional control variable, and employ weighted least squares regression (although with different weighting variables). There are major differences: (1) we use a super-size dummy from a natural experiment while Bedard and Kuhn use observed class size either as a cubic or as a series of class size binaries and (2) ours is a small private university with no section larger than 45 students while theirs is a public university with sections that can exceed 300 students. Our estimated impact of super-sizing (a conditional increase of approximately 13.4 students or from roughly 25 to 38 students) is to lower an instructor rating by 0.111 points. Applying Bedard and Kuhn's estimated coefficients to those class sizes lowers instructor effectiveness ratings by 0.078 points in the cubic variant or by 0.166 points in the variant using class size binaries. That our estimate falls between their estimates lends some credence to the idea that reasonable estimates of class size impacts can be obtained without controlling for potential endogeneity in the assignment.

While it is clear from these results that students generally rate courses and faculty in large sections less favorably than smaller sections, the above results do not clarify why that is the case. We address this in Table 3 by estimating the impacts of class size on specific course attributes and faculty practices. In each case, all of the above control variables, including time, faculty, and course fixed effects are also included in the regressions (but the results are not shown). We present only the results of class size on the course attributes listed down the left hand side of Table 3.[3] We find that class size has a negative and statistically

[3] Weighted 2SLS regressions were estimated in the same manner as they were for Table 2. All regressions in Table 3 are based on 1,928 sections/observations with the exception of workload in the course. Because this question changed in the 2008 course evaluations, only the 1,794 sections prior to 2008 are used in the workload regression.

DOI: 10.2202/1935-1682.2803

significant impact on student assessment of eight faculty practices. At the 99% level, students in larger sections gave lower ratings to the clarity of presentations; at the 95% level to the effectiveness of the teaching methods, daily preparedness of the instructor, and adequacy of graded material relative to course content; and at the 90% level to the instructor's effectiveness in stimulating interest, enthusiasm for teaching the class, timeliness of feedback on assignments, and usefulness of the text. At the 90% level, students perceived the pace of the course to be faster in larger sections. Several other attributes were rated lower in larger sections, such as the amount of critical thinking, the availability of the instructor to the students, the respect the instructor had for the students, and timeliness of feedback, although none of these were statistically significantly different from zero, at conventional levels.

Clearly, class size is perceived by students as significantly altering many important aspects of their courses. These results suggest that the negative influence of class size on course outcomes works through altering the attributes of the instructors' courses. Faculty that handle large sections appear to change their courses in ways that negatively affect course outcomes.

Table 3
Regression Results
Impact of Class Size on Course Attributes
Dependent variables are in the rows

Course or Faculty Attribute	Impact of Class Size
Course Workload	0.0002 (0.0024
Course Level of Difficulty	0.0034 (0.0022)
Level of Critical & Analytical Thinking	-0.0016 (0.0021)
Pace of Course	0.0024 * (0.0013)
Clear and Understandable Presentation	-0.0086 *** (0.0033)
Effectiveness of Teaching Methods	-0.0077 ** (0.0034)
Instructor's Daily Preparation for Class	-0.0050 ** (0.0022)
Effectiveness in Stimulating Interest	-0.0063 * (0.0033)
Enthusiasm for Teaching Course	-0.0041 * (0.0023)
Availability Outside the Classroom	-0.0024 (0.0027)
Respect for Students in the Class	-0.0034 (0.0027)
Graded Material vs. Course Content	-0.0079 ** (0.0033)
Adequacy of Comments on Student Work	-0.0056 * (0.0030)
Timeliness of Feedback on Student Work	-0.0020 (0.0032)
Usefulness of Text, etc. to Learning	-0.0060 * (0.0034)

Notes: All of the control variables from Table 2, including faculty, semester and course fixed effects, were included in the regressions but not in this table. Standard errors are shown in parentheses; * indicates two-tailed significance at the 90% level, ** at 95%, and *** at 99%.

DOI: 10.2202/1935-1682.2803

V. Conclusion

The evidence found in this analysis unequivocally leads to the conclusion that class size has a negative impact on the student-rated outcomes of amount learned, instructor rating, and course rating. This negative relationship between class size and student-rated outcomes is found utilizing a natural experiment in class-size, conditional on faculty, course, and time fixed effects. These results corroborate the negative relationship found in previous studies by demonstrating that the negative relationship is not attributable solely to endogenous variation in class size across instructors. Rather, controlling for the instructor, course, and numerous other characteristics, larger class sizes resulting from an exogenous policy change led to lower student ratings, an effect that was not eliminated with more experience teaching the larger sections.

Additionally, the analysis above reveals that class size primarily influences student-rated outcomes by altering certain aspects of courses that students find beneficial and helpful in learning. For example, large class sizes are correlated with less clarity in class presentations, less preparation, less enthusiasm, lower effectiveness in stimulating interest, less effective teaching methods, less adequate graded material, slower return of assignments, and less useful course materials. It is reasonable to assume that these course and instructor attributes are positively related to students' overall course assessments.

Reducing class sizes will lead to significant improvements in student ratings and self-reported course outcomes. Course attributes, such as clarity of presentation, instructor preparedness, and stimulating student interest, important to student learning and how much a student gets out of a course suffer when class size is increased. On the other hand, our results also suggest that student load (the total number of students taught by a faculty member in a semester) may also be important in determining course outcomes. We would caution against administrative policies of hiring faculty to teach numerous sections of a course in order to minimize class sizes due to the potential importance of student load in determining course outcomes. Policies of hiring faculty to teach numerous small sections, in order to reduce class sizes with the aim of faring better in institutional rankings, should be weighed against the impact these policies may have on student load and the important impact it may play in student-rated outcomes.

References

Achilles, Charles, Harman, Patrick, and Egelson, Paula (1995). "Using Research Results on Class Size to Improve Pupil Achievement Outcomes." *Research in Schools*, *2*(2), 23-30.

Akerhielm, Karen (1995). "Does Class Size Matter?" *Economics of Education Review*, *14*(3), 229-41.

Angrist, Joshua, and Lavy, Victor (1999). "Using Maimondides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, *114*, 533-75.

Averett, Susan, and McLennan, Michele (2004). "Exploring the Effect of Class Size on Student Achievement : What Have We Learned over the Past Two Decades?" In Geraint Johnes and Jill Johnes (Eds.), *International Handbook on the Economics of Education* (pp. 329-67). Cheltenham, UK: Edward Elgar Publishing.

Bandiera, Oriana, Larcinese, Valentino, and Rasul, Imran (2010). **"**Heterogeneous Class Size Effects: New Evidence from a Panel of University Students." *Economic Journal*, *120*(549), 1365-98.

Bedard, Kelly, and Kuhn, Peter (2008). "Where Class Size Really Matters: Class Size and Student Ratings of Instructor Effectiveness." *Economics of Education Review*, *27*(3), 253-65.

Correa, Hector (1993). "An Economic Analysis of Class Size and Achievement in Education." *Education Economics*, *1*(2), 129-35.

Dickens, William (1990). "Error Components in Grouped Data: Is It Ever Worth Weighting?" *The Review of Economics and Statistics*, *72*(2), 328-33.

Goldhaber, Daniel, and Brewer, Dominick (1997). "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources*, *32*(3), 505-23.

Hoxby, Caroline (2000). "The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation." *Quarterly Journal of Economics*, *115*(4), 1239-85.

DOI: 10.2202/1935-1682.2803

Kokkelenberg, Edward, Dillon, Michael, and Christy, Sean (2008). “The Effects of Class Size on Student Grades at a Public University.” *Economics of Education Review*, *27*(2), 221-33.

Krueger, Alan (1999). “Experimental Estimates of Education Production Functions.” *Quarterly Journal of Economics*, *114*(2), 497-532.

Kulik, James (2001). “Student Ratings: Validity, Utility, and Controversy.” *New Directions for Institutional Research*, *109*, 9-25.

Lazear, Edward (2001). “Educational Production.” *Quarterly Journal of Economics*, *116*(3), 777-803.

Walia, Bhavneet (2008). *Three Essays in Health and Labor Economics* Ph.D. Dissertation. Kansas State University.

Westerlund, Joakim (2008). “Class Size and Student Evaluations in Sweden.” *Education Economics*, *16*(1), 19-28.