Does University Class Size Matter? Evidence from Course Micro Data

Xiufeng Xing*
Dept. of Economics, Business School
Qingdao University of Technology
Qingdao, China

Aiqin Xi
Dept. of Economics, Business School
Qingdao University of Technology
Qingdao, China

Xuli Zhao
Office of Teaching Affairs
Qingdao University of Technology
Qingdao, China

Abstract—Using the data from macroeconomics test scores and students’ basic information, we estimate class size and average test scores for undergraduate students in Qingdao University of Technology. We find a substantial class size effect. That is, given the two different class sizes—the medium size and the large size, the test score in the medium class size is two scores higher than that of the latter, so to speak, the smaller class size effect is more pronounced. Our finding is consistent with those of others concerning the relationship between class size effect and test scores. Incidentally, science major might enhance score compared to economics major. For policy maker and university administration, this is especially meaningful in terms of the match of classroom resources and students’ course enrollment, as well as teacher assignment or hiring.

Keywords—class size; final test scores; Chinese ordinary universities; macroeconomics; micro data

I. INTRODUCTION

Class size is a critical element influencing students’ learning outcomes and therefore, takes researchers’ attention. A number of viewpoints have been presented in the literature to explore the effects of different class size. Galton (1998) kicks off the research on class size effect. The topic of class size reduction (CSR) has become the heated debate ever since then. Blatchford, P. and Russell, A. summarize the associated aspects including class size and students learning outcome, class size and classroom process, as well as class size and effective learning. Our comments will be brief and references to the existing literature will be less exhaustive, we only focus on those relative to this issue, namely, center on the relation between final test score and class size in universities. First and for most, much empirical research on this subject typically takes final course grades, standardize test scores, or course evaluation data as the outcome measures. Quite a number of international studies show positive impact of small class size on student performance and satisfaction, while others do not. Claudio and Gaston (2016) find that there exists a negative effect of increasing class size by one standard deviation of roughly 0.187 SDs of their outcome measure. Angrist and lavy (1999) and Krueger (2003) both find positive effects of small class size. Johnson (2010) explored the effect of class size on undergraduate students’ achievements at a public university from a broad len of class size among various disciplines. In the analysis, the author points out that small size classes can provide students with many opportunities to participate in classes and have positive peer effects. The author also demonstrates that the effect of class size on students’ final grades diminishes as the class size increases. Interestingly, even if capitalizing on the same data set, researchers may arrive at dissimilar outcomes. For instance, although based on the identical data from the third edition of the Test of Understanding in College Economics (TUCE III), Lopus and Maxwell (1995), Zietz and Cochran (1997), and Kennedy and Siegfried (1997) get different conclusions. While Lopus and Maxwell (1995) find a positive correlation between class sizes and test scores, Zietz and Cochran (1997) show that increasing class size over 30 students has a negative impact on test results. As for Kennedy and Siegfried (1997), they find no effect of class size on test performance.

Second, when it comes to definition/specification of the class size, there is no consensus. Overall, some studies employ different definitions of small and large class, others set a limit to only a few classes. To name a few only, Hancock (1996) examines the effect of class size on student grades in sections which contain 39 students on average and are defined as “normal”, and sections with 118 students on average which are referred to as “megasections”. Borden and Burton (1999) defines a class as “small” if it comprises 5–30 students, as “medium” if it comprises 31–90 students, and as “large” if it covers more than 90 students. These differences in class sizes under consideration point out another source of discrepancies in research outcomes. Raimondo et al. (1990) differentiate between large (200–350 students) class and small (25–35 students) one. For Scheck et al. (1994) small classes have a
maximum of 38 students. Gibbs et al. (1996) refers to class as “small” if it contains equal to or less than 30 students and “large” if it contains over 70 students. In our paper, we define a class as “medium” if it contains 50-55 students, as “large” if it comprises 90+ students.

Last but not least, regarding the causal link between class size and student test scores, there are several methods, such as linear, logarithmic, fuzzy regression-discontinuity design, etc. we use the OLS regression to estimate the causal link of class size on student test score and major.

The goal of this paper is to contribute to better understanding of the causal links between class size and test scores and hence learning outcomes of undergraduate students in Chinese ordinary universities like Qingdao University of Technology. The paper focuses on these issues by heavily relying on micro data from first-hand macroeconomics course and student information. The resulting conclusion tends to help Chinese ordinary universities policy-makers better allocate the resources of students and classroom and professional teachers in future.

The remainder of the paper proceeds as follows: Section two describes the background and course micro data. Section three discusses the methodology used. The empirical results are reported in Section four and Section five concludes the paper.

II. BACKGROUND AND COURSE DATA

In the fall semester of 2018-2019 academic year, on the order of 800 undergraduates in business school of Qingdao University of Technology chose macroeconomics, and six one-time teachers lectured in this course. That is, the teachers have classes at same time but in different classrooms, each lecturer is assigned to either a medium class size or/and a large one, respectively. At the end of the semester, the said 800 students take the final exam simultaneously with the identical test paper, and their papers are collaboratively examined by the six lecturers after the test, namely, each lecturer is assigned to check one type of test question and offers grades to all the test takers to make sure the same scale of grade lenience. Our research can be considered as a quasi-natural experiment in a sense, so to speak.

Our paper comprises two undergraduate classes, with one class consisting of 52 students, while another 101. At the end of the semester, except three of them requested for delayed tests, 150 students in the two classes attended the same final exam. Based on the first-hand score microdata, as well as other information associated with the above 150 students, we make a short panel data or a cross section data, so to speak, since the course data covers only one time period, and make analysis accordingly.

Our course data comes from Macroeconomics taught at Qingdao University of Technology between the first semester of 2018 and 2019. Overall, the data consists of 1 semester, one course, and 1 teacher, for a total of 150 observations. The data is a short panel one, since at the end of every semester, the students are required to choose course online for the next semester, note that under Qingdao University of Technology’s new course-choosing policy, students are no longer select the same teacher and complete both the microeconomics and macroeconomics as before. In this case, we can only collect one semester data relative to the course of macroeconomics. For a regular class size in most Chinese universities, the number of students of on the order of 30. Currently, the regular class size in Qingdao University of Technology contains around 30-36 students in line with per major, and each professional teacher is randomly assigned to teach course(s) relative to his or her major. Regarding macroeconomics, each teacher instructs at most two different class sizes, with one class size of roughly 50 students and about 100 for large class size allow for classroom space as well as the actual number of students course choosing situation. Additionally, the macroeconomics and microeconomics can be chosen across all majors in Qingdao University of Technology which was implemented from 2018. As for the class size in our case, it is convenient to define the class with 52 students enrolled as “medium”, and “large” for the other class with 101 students enrolled.

III. METHODS

As to the above two classes: medium size class, which contains 51 exam takers, has the average score of nearly 75, while the larger one simply 73. We therefore would like to ask if class size matters in determining students’ average test score. with everything else being equal, our case seems like a quasi-natural experiment, it is natural and convenient for analyzing the relationship among the variables of test score, class size, as well as other factors like major. We show great interest in this problem since undergraduates face the same scenario: the same one time teacher and learning resources. To explore this, we first set up the following empirical models:

\[
\text{Log}\text{score} = \beta_0 + \beta_1 \text{size} + \beta_2 \text{major} + \beta_3 \text{gender} + u
\]  

\[
\text{Log}\text{score} = \beta_0 + \beta_1 \text{size} + \beta_2 \text{major} + v
\]  

\[
\text{Log}\text{score} = \beta_0 + \beta_1 \text{size} + \beta_2 \text{major} + \beta_3 \text{size} \times \text{major} + w
\]

Where u, v, w denotes the corresponding error term in the above equations. The null assumption is H0: causal link between class size and test scores, the alternative assumption is HA: No causal link between class size and test scores. Intuitively speaking, the gender variable in (1) makes sense, because in the effect, the female undergraduate students in our economics major outnumber the male ones, and this is usually the case in many Chinese universities like Qingdao University of Technology. Moreover, the regression verifies this assumption. However, because we primarily want to examine whether or not there exist a causal link between the variables of test score and class size, in this case, we partial out the gender effect in (1) and mainly focus on (2). By construction, for the sake of simplicities, we omit interaction term(s) in the first two equations. As far as (3) is concerned, the interaction term means a mixed effect of size and major. Further, the signs of the interaction term may be different: if a minus sign appears, it probably means that the interaction of large class size and non science major reduces the average score, and vice versa.

Alternatively, we might explore the relationships among test score (in 100 points), class size as well as major, only to find the regression outcome shows that science major plays a
decisive role in the distribution of final test score, which is counterfactual in our case, since science major undergraduate students only take quite a small portion, moreover, the individual score in science major are not quite high per se. That is why we ignore this fashion.

IV. EMPIRICAL RESULTS

![Fig. 1. Test Scores of Two Different Class Sizes](image)

One can see from the above picture (Fig. 1) that, overall, the score distribution in the medium size class is less fluctuation than that of the large size one. To be more specific, besides the average test scores difference, on the order of 7.8 percent exam takers in the medium size class fail the final test, by contrast, approximately 15.2 percent exam takers in large size class fail to pass the test. Table 1 spells out the details of associated dependent variables and independent variables.

| Variable | Description | Obs | Mean  | Std dev | Min  | Max  |
|----------|-------------|-----|-------|---------|------|------|
| score    | score of full sample | 150 | 73.18 | 13.17   | 11   | 92   |
| score1   | score of medium class | 51  | 74.73 | 11.20   | 22   | 89   |
| score2   | score of large class | 99  | 72.38 | 14.07   | 11   | 92   |
| size     | 1 if medium class; 0 otherwise | 0.34 | 0.48  |         |      |      |
| major    | 1 if science major; 0 otherwise | 0.19 | 0.39  |         |      |      |

One can see from TABLE I, the dependent variables include score, score1 and score 2, while independent variables are size and major. The average score of students of full sample is 73.2, while for two subsamples in medium class and in large size class, the average score for students is about 75 and 72, respectively. It is obvious that the average score of the large size class is 0.8 lower than that of the average score of full sample. By contrast, the average score of the medium size class is approximately 1.6 higher than that of full sample.

TABLE II. ESTIMATES OF CLASS SIZE ON TEST SCORE, CONTROLLING FOR MAJOR.

| OLS estimates of class size on test score, controlling for major |
|------------------------------------------------------------------|
| (1)                                                              |
| size                                                             |
| major                                                            |
| Observations                                                    |

* Notes: This table shows results of OLS regression of score on class size, with and without including major as a control. Heteroscedasticity-robust standard errors are presented in t parenthesis. ***, ** and * indicate the 0.01, 0.05 and 0.1 significant level, respectively.

TABLE II displays the causal relations among scores, class size and major. When logscore is regressed on class size without including major as a control, the class size effect is more pronounced. While regressing with major, the outcome is less obvious. With p-value being 0.443, it is insignificant such that we should accept the null hypothesis: there is causal link between class sizes and test scores. Further, the Stata output gives the heteroskedasticity-robust F statistic. One can see from the t statistics that these variables are going to be jointly insignificant. The F test verifies this, with p-value = 0.601. Therefore, we should accept the null assumption at 5% significant level. In other words, the test score is negatively correlated with class size: given the two class sizes, the smaller one has the higher the test score. Therefore, medium size seems to be more conducive to high final average test score. Student in medium size class appears to have 4.7 percent higher average score points than that of large class size. Incidentally, science major might enhance score compared to economics major. To be specific, science major undergraduate students can improve 0.8 percent score points than that of economics or similar majors.

Also, we allow for student gender in our analysis, as a comparison with (2), the coefficients of (1) is list in TABLE III:

TABLE III. ESTIMATES OF CLASS SIZE, MAJOR AND GENDER ON TEST SCORE

| Score | Coef. | Std. Err. | t     | [95% Conf. Interval] |
|-------|-------|-----------|-------|----------------------|
| size  | 0.0537| 0.055     | 0.98  | -0.055               | 0.162 |
| major | 0.089 | 0.071     | 1.25  | 0.213                | -0.052 | 0.231 |
| female| 0.153 | 0.049     | 3.11  | 0.002                | 0.056 | 0.249 |
| cons  | 4.143 | 0.044     | 95.22 | 0.000                | 4.057 | 4.229 |

From TABLE III, one can see that gender is significant both statistically and economically, with the p value being 0.002. Interestingly, class size becomes the least statistically significant variable among the three independent variables. And to account for this we will not be discussed here.

Finally, when we take the interaction term in (3) into consideration, and have done the resulting regression, only to find that there is actually a minus sign for the coefficients of interaction term, with the value of -0.154. Put it differently, it indicates that the interaction of large class size and non science major lowers the average test score. Therefore, in either case, the conclusion that smaller class size is superior to larger size class in term of students learning outcome holds.
V. Conclusions and Policy Implications

A. Conclusions

In our study, ceteris paribus, we arrive at that, medium class size contributes more to students’ average test score which is mostly consistent with many similar research perspectives mentioned earlier in this paper.

1) It is likely that a causal link between class size and test scores of macroeconomics exists: student in a smaller class size has a relatively higher average test scores.

It indicates that the average score of students in small university class size is around 2.3 points higher than that of larger one’s. Furthermore, the course mean of large class size is 1.5 points lower than the average mean which is 73.2 of the full sample (150 students). This outcome may be due to large classes provide students with less engagements and peer interactions (Johnson, 2010).

2) Science major may enhance test score compared to economics major. This is intuitively reasonable since science majors have a comparatively more solid foundation in terms of the comprehension of calculus or/and probabilities theory as well as other science disciplines. It is true that economics relates closely with mathematics knowledge.

Beyond the said conclusions, there might be some indirect negative effects of class size on students’ learning outcomes which we are unaware of and should become a future study issue to be delved into.

B. Policy Implications

Why does the variation of class size effect buy us? It is mostly because the correlation between university class size and students’ learning outcome is of huge importance for university administers, policy makers, classroom teachers as well as students. For university administers and policy makers, there is a tradeoff between open a new class or enlarge the current class size should the student enrollment exceed a certain number given the limited classroom size, which might need to assign or hire new teachers to give lectures accordingly. In addition, the macroeconomics is open to various majors, due to the different prior academic background, students may choose this course either as core or selective course based upon his or her demand. Plus, university administrators should allow courses like macroeconomics and microeconomics, due to the integrality of this kind of courses, the two courses should have been chosen together such that it is convenient and wise for students to choose these courses at a time. For now, however, under the current course option policy in Qingdao University of Technology, students have to choose each of the two courses in different semester which is less flexible in terms of course option design.

As for classroom teachers, they are likely to take dissimilar methodologies to address the situations, that is, for a large size class, the teacher may take on lectures as a main means to enhance the learning efficiency; given a smaller class size, however, the professional teacher can take a more flexible approach to help undergraduate students boost their learning outcome, such as lectures couple with class discussion. In a word, a classroom teacher should adapt to difference situations and develop the ability of learning by teaching such that to maximize the learning outcome for undergraduate students. With regard to students, if they are randomly assigned to a bigger size class, they may suffer a little bit from the less peer interactions and a uniformed teaching style adopted by the lecturer. In that case, undergraduate students need to take initiatives to make up for the defects individually.

Finally, a limitation of this paper should be acknowledged. In particular, due to the lack of information on undergraduate students’ macroeconomics across Chinese ordinary universities, the said conclusion may be weakened by the small sample size. Beyond that, to enhance the accuracy of estimation, more variables like students in-class attendance, teacher experience and qualification should also be considered in future research.

REFERENCES

[1] Galton, M., “Class size: A critical comment on the research,” International Journal of Educational Research, 1998, pp.809-818.
[2] Blatchford, P., and Russell, A., “New ways of thinking about research on class size: an international perspective,” International Journal of Educational Research, in press.
[3] Claudio Sapelli, and Gaston Ilanes, “Class size and teacher effects in higher education,” Economics of Education Review, 2016, pp. 19-28.
[4] Bekir Celika, and Nizamettin Koca, “Effect of Class Size on University Entrance Exam in Turkey”, Procedia - Social and Behavioral Sciences, 2015, pp. 191, 919 – 924.
[5] Angrist, J. D., and Lavy, V., “Using maimonides’ rule to estimate the effect of class size on scholastic achievements”, The Quarterly Jour- nal of Economics, 1999, pp. 533–575.
[6] Krueger, A. B., “Economic considerations and class size”, Economic Journal, vol.113 (485), 2003, pp. F34–F63.
[7] Johnson, I. Y., “Class Size and Student Performance at a Public Research University: A Cross-Classified Model”, Research in Higher Education, 2010, pp. 701–723.
[8] Lopus, J. S., and Maxwell, N. L., “Should we teach microeconomic principles before macroeconomic principles?” Economic Inquiry, 1995, pp.336–350.
[9] Zietz, J., and Cochran, H. H., “Containing cost without sacrificing achievement: Some evidence from college-level economics classes”, Journal of Education Finance, 1997, pp.177–192.
[10] Kennedy, P. E., and Siegfried, J. J., “Class size and achievement in introductory economics: Evidence from the TUCE III data”, Economics of Education Review, 1997, pp. 385–394.
[11] Hancock, T. M., “Effects of class size on college student achievement”, College Student Journal, 1996, pp. 479–481.
[12] Borden, V. M. H., and Burton, K. L., “The impact of class size on student performance in introductory courses”, Paper presented at the 39th annual conference of the AIR in Seattle, WA, 1999.
[13] Raimondo, H. J., Esposito, L., and Gershenberg, I., “Introductory class size and student performance in intermediate theory courses”, Journal of Economic Education, 1999, pp.369–381.
[14] Schecter, C., Kinicki, A., and Webster, J., “The effect of class size on student performance: Development and assessment of a process model”, Journal of Education for Business, 1994, pp.104–111.
[15] Gibbs, G., Lucas, L., and Simonite, V., “Class size and student performance: 1984–94”, Studies in Higher Education, 1996, pp.261–273.