Economics of Education Review 52 (2016) 19–28

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Class size and teacher effects in higher education

Claudio Sapelli[a,*], Gastón Illanes[b]

[a] Economics Department, Pontificia Universidad Catolica de Chile, Chile
[b] Economics Department, MIT, USA

ARTICLE INFO

Article history:
Received 23 October 2013
Revised 3 November 2015
Accepted 4 January 2016
Available online 24 February 2016

JEL Classification:
I21
I23
I28

Keywords:
Class size
Teacher effects
Student evaluations

ABSTRACT

Using student evaluations of their instructor as an outcome measure, we estimate and compare class size and teacher effects for higher education, with an emphasis on determining whether a comprehensive class size reduction policy that draws on the hiring of new teachers is likely to improve educational outcomes. We find that first time teachers perform significantly worse than their peers, and we find substantial class size effects. Hence higher education institutions face a tradeoff if they wish to increase admission. This tradeoff implies that as class size increases, at first the negative class size effect is smaller than that of introducing a first time teacher. However, beyond a certain level, the class size effect dominates and it is better to create a new class with a first time teacher.[1]

1. Introduction

Several studies have estimated the effect of class size on learning outcomes, highlighting that smaller classes foster learning. However, when recommending smaller classes as a policy, it is often forgotten that the teachers hired to work in those classes may not be of the same quality as those currently teaching. Thus, the effect of reducing class size on outcomes will depend crucially on the balance between the positive effect of a smaller class and the potentially negative effect of the quality gap between inframarginal and marginal teachers. This work gives insights for higher education on the decision whether to increase class size with existing teachers or hire new (first time) teachers. We provide evidence on class size and first time teacher effects, using teacher evaluation surveys from the Economics Department and the Business School at Pontificia Universidad Católica of Chile (FACEAPUC). First time teacher effects are relevant for this discussion because the most likely avenue for an increase in the number of teachers in Higher Education is hiring first time teachers.

We use student evaluation data as an outcome measure. It can be thought of as an indicator of student learning or an indicator of student satisfaction. Although interpreting student evaluations as an indicator of learning has its problems (see Braga, Paccagnella, & Pellizzari, 2014; Carrell & West, 2010), this method also has distinct advantages over other output measures for evaluating teachers, such as test scores. Hanushek (2003) and Krueger (2003) argue that estimating the effect of class size on learning using test scores raises major concerns, since results are sensitive to the econometric specification used and to the outcome variable in question. Also, there is research linking student satisfaction to effective learning (Theall & Franklin, 2001),

* Corresponding author at: Vicuña Mackenna 4860, Santiago, Chile. Tel.: +562 23544003.
E-mail addresses: csapelli@uc.cl (C. Sapelli), gastonillanes@gmail.com (G. Illanes).

[1] We would like to thank Matías Covarrubias and Fernanda Rojas for excellent research assistance. We would also like to thank comments received in the internal workshop of Pontificia Universidad Católica de Chile's Economics Department and in the 2011 Yearly Congress of the Economics Society of Chile. We also thank two anonymous referees whose advice greatly improved the paper. The usual disclaimer applies.

http://dx.doi.org/10.1016/j.econedurev.2016.01.001
0272-7757/

and research on student evaluations that provides evidence that student ratings are reliable, valid, unbiased, and useful (Murray, 1994). Finally, Bedard and Kuhn (2008) build on this, arguing that student evaluations are better indicators of student learning. We build on this research by using a FRDD methodology to identify causal links, taking advantage of a discontinuity in class size we observe in our data.

We find that there is a negative effect of increasing class size by one standard deviation of roughly 0.187 SDs of our outcome measure. This is similar in size to the lower bound of those found in the literature (Hanushek & Rivkin, 2010). We also find that the average impact of a first time teacher is −0.41 standard deviations. That is, a first time teacher is substantially worse than infra-marginal teachers. As we will show, there is also substantial risk in hiring a new teacher. In higher education, where teaching loads for full time professors are not flexible, administrators often face the decision of increasing class size or hiring a first time teacher. We find that both choices entail a drop in student satisfaction, and hence that the decision rule would imply increasing the class size up to a certain level and then splitting the class and hiring a first time teacher. We give evidence on the magnitude of these effects and discuss how we infer decisions are taken in this context, particularly considering that administrators face uncertainty.

The paper proceeds as follows: Section 2 summarizes the relevant literature on the education production function and on student evaluations, Section 3 presents our data, Section 4 explains the econometric methodology used, Section 5 presents our results, and Section 6 concludes.

2. Literature review

Studies that estimate teachers' effects on achievement using longitudinal data, such as Rockoff (2004), have become a first step in solving many puzzles in the production function of achievement. Estimates suggest that the best teacher may raise achievement by as much as half a standard deviation. Though this literature also finds that credentials do not explain teacher effects for the most part, the exception is that very inexperienced teachers have worse effects, and that the effects of increased experience plateau after four to five years.

This finding has led to the need to measure teacher effects and class size effects and trade off one against the other. If we are to go by the median estimate in the literature then teacher effects are between two times and six times larger than class size effects. Though results in the literature vary with methodology and data set (see Meghir and Rivkin, 2011 for a thorough treatment), there is an emerging consensus regarding the great heterogeneity of teacher quality and its importance. It is in this area of the literature that we wish to contribute.

The most influential studies of class size reduction are those based on the Student Teacher Achievement Ratio, or STAR, a study conducted in Tennessee in the late 1980s. Among them possibly Krueger's (2003) analysis is the most cited one. He finds that elementary school students randomly assigned to small classes outperformed their classmates assigned to regular classes by about 0.22 standard deviations after four years. Other credible studies that also find positive effects of class size reduction find smaller effects. For example, Rivkin, Hanushek, and Kain (2005) examine the effects of natural variation in class size in Texas in the mid-1990s. The estimated effects were about half the size of the effects found in Krueger (2003). International studies also provide positive evidence for the effects of class-size reduction. Angrist and Lavy (1999) take advantage of a class-size limit in Israel of 40 students. They find positive effects of smaller classes, with effect sizes that are on the lower end of those found in the STAR study. Jepsen and Rivkin (2009) examine the class size reduction enacted in 1996 in California. The program was designed to reduce class size by ten students per class, from 30 to 20. They also find positive effects for class-size reduction that are about half as large as those found in Tennessee. Interestingly Jepsen & Rivkin (2009) study also the changes in the teachers required by this change. They find that increases in the numbers of new and not-fully-certified teachers offset much of these gains. In other words, students who ended up in the classrooms of teachers new to their classrooms and grades suffered academically from the teacher's inexperience by almost the same amount as they benefited from being in a smaller class. Summarizing, it appears that large class-size reductions, on the order of magnitude of 7–10 fewer students per class, can have important long-term effects on student achievement. The largest estimates of the magnitude of class-size effects are those produced by Krueger (1999), who found that the students in classes that were 7 to 8 students smaller on average than regular-sized classes performed about 0.22 standard deviations better on a standardized test. This means that students performed about 3 percent of a standard deviation better for every 1 student less in the class.This leads to think that if there is a reduction of 10 students, the effect will be of 0.30 standard deviations. Since most other studies find results that are about half of these (or somewhat lower than that) this has led (Hanushek & Rivkin, 2010) to argue that the literature shows that the effect of a ten student reduction in class size is between 0.10 and 0.30 standard deviations of the dependent variable. At the postsecondary level, Bedard and Kuhn (2008) argue that student evaluations may be a useful indicator of a teacher' s performance. Relative to this work, we tackle the problem with an identification strategy that better deals with endogeneity in class size.

There is value in using student ratings for teacher evaluation. Cashin (1999) performs a meta-analysis of the research and concludes that "student ratings tend to be statistically reliable, valid and relatively free from bias or need for control; probably more so than other data used for evaluation" . There is, however, no consensus regarding the adequacy of student ratings as a measure of instructor or course effectiveness. Be that as it may, they are indicators of student satisfaction (Theall & Franklin, 2001). Moreover, there are positive and significant correlations between student ratings and student learning; and between student ratings and observer, peer and alumni ratings (Greenwald, 1997; McKeachie, 1997). However, there are several drawbacks to using student evaluations as an outcome measures. There is controversy regarding the correlation