

Evaluating students' evaluations of professors[☆]Michela Braga^a, Marco Paccagnella^b, Michele Pellizzari^{c,*}^a Bocconi University, Department of Economics, Italy^b Bank of Italy, Trento Branch, Italy^c University of Geneva, Institute of Economics and Econometrics, Switzerland

ARTICLE INFO

Article history:

Received 2 August 2013

Received in revised form 22 April 2014

Accepted 23 April 2014

Available online 5 May 2014

JEL classification:

I20

M55

Keywords:

Teacher quality

Postsecondary education

Students' evaluations

ABSTRACT

This paper contrasts measures of teacher effectiveness with the students' evaluations for the same teachers using administrative data from Bocconi University. The effectiveness measures are estimated by comparing the performance in follow-on coursework of students who are randomly assigned to teachers. We find that teacher quality matters substantially and that our measure of effectiveness is negatively correlated with the students' evaluations of professors. A simple theory rationalizes this result under the assumption that students evaluate professors based on their realized utility, an assumption that is supported by additional evidence that the evaluations respond to meteorological conditions.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The use of anonymous students' evaluations of professors to measure teachers' performance has become extremely popular in many universities (Becker & Watts, 1999). They normally include questions about the clarity of lectures, the logistics of the course, and many others. They are either administered during a teaching session toward the end of the term or, more recently, filled on-line.

The university administration uses such evaluations to solve the agency problems related to the selection and motivation of teachers, in a context in which neither the types of teachers, nor their effort, can be observed precisely. In fact, students' evaluations are often used to inform hiring and promotion decisions (Becker & Watts, 1999) and, in institutions that put a strong emphasis on research, to avoid strategic behavior in the allocation of time or effort between teaching and research activities (Brown & Saks, 1987; De Philippis, 2013).¹

[☆] We would like to thank Bocconi University for granting access to its administrative archives for this project. In particular, the following persons provided invaluable and generous help: Giacomo Carrai, Mariele Chirulli, Mariapia Chisari, Alessandro Ciarlo, Alessandra Gadioli, Roberto Grassi, Enrica Greggio, Gabriella Maggioni, Erika Palazzo, Giovanni Pavese, Cherubino Profeta, Alessandra Startari and Mariangela Vago. We are also indebted to Tito Boeri, Giovanni Bruno, Giacomo De Giorgi, Marco Leonardi, Vincenzo Mariani, Tommaso Monacelli, Tommy Murphy and Tommaso Nannicini for their precious comments. We would also like to thank seminar participants at the Bank of Italy, Bocconi University, International Workshop on Applied Economics of Education, London School of Economics, UC Berkeley, Università Statale di Milano and LUISS University. Davide Malacrino and Alessandro Ferrari provided excellent research assistance. Michele Pellizzari is also affiliated to IZA, CREAM, FRDB and NCCR-LIVES. The views expressed in this paper are solely those of the authors and do not involve the responsibility of the Bank of Italy. The usual disclaimer applies.

* Corresponding author at: University of Geneva, Department of Economics, Uni-Mail, 40 Bd. du Pont d'Arve, CH-1211 Geneva 4, Switzerland. Tel.: +41 22 37 98 915.

E-mail addresses: michele.pellizzari@unige.ch (M. Pellizzari).

¹ Although there is some evidence that a more research oriented faculty also improve academic and labor market outcomes of graduate students (Hogan, 1981).

The validity of anonymous students' evaluations rests on the assumption that, by attending lectures, students observe the ability of the teachers and that they report it truthfully when asked. While this view is certainly plausible, there are also many reasons to question the appropriateness of such a measure. For example, the students' objectives might be different from those of the principal, i.e. the university administration. Students may simply care about their grades, whereas the university cares about their learning and the two might not be perfectly correlated, especially when the same professor is engaged both in teaching and in grading. Consistent with this interpretation, [Krautmann and Sander \(1999\)](#) show that, conditional on learning, teachers who give higher grades also receive better evaluations. This finding is confirmed by several other studies and is thought to be a key cause of grade inflation ([Carrell & West, 2010](#); [Johnson, 2003](#); [Weinberg, Fleisher, & Hashimoto, 2009](#)).

Measuring teaching quality is complicated also because the most common observable teachers' characteristics, such as qualifications or experience, appear to be relatively unimportant ([Hanushek & Rivkin, 2006](#); [Krueger, 1999](#); [Rivkin, Hanushek, & Kain, 2005](#)). Despite such difficulties, there is evidence that teachers' quality matters substantially in determining students' achievement ([Carrell & West, 2010](#); [Rivkin et al., 2005](#)) and that teachers respond to incentives ([Duflo, Hanna, & Ryan, 2012](#); [Figlio & Kenny, 2007](#); [Lavy, 2009](#)). Hence, understanding how professors should be monitored and incentivized is essential for education policy.

In this paper we evaluate the content of the students' evaluations by contrasting them with objective measures of teacher effectiveness. We construct such measures by comparing the performance in subsequent coursework of students who are randomly allocated to different teachers in their compulsory courses. We use data about one cohort of students at Bocconi University – the 1998/1999 freshmen – who were required to take a fixed sequence of compulsory courses and who were randomly allocated to a set of teachers for each of such courses.

We find that, even in a setting where the syllabuses are fixed and all teachers in the same course present exactly the same material, professors still matter substantially. The average difference in subsequent performance between students assigned to the best and worst teacher (on the effectiveness scale) is approximately 23% of a standard deviation in the distribution of exam grades, corresponding to about 3% of the average grade. Moreover, our measure of teaching quality is negatively correlated with the students' evaluations of the professors: teachers who are associated with better subsequent performance receive worst evaluations from their students. On the other hand, teachers who are associated with high grades in their own exams rank higher in the students' evaluations.

These results question the idea that students observe the ability of the teacher during the class and report it (truthfully) in their evaluations. In order to rationalize our findings it is useful to think of good teachers – i.e. those who provide their students with knowledge that is useful in future learning – as teachers who require effort from their students. Students dislike exerting effort, especially the least able ones, and when asked to evaluate the teacher they do so on the basis of how much they enjoyed the course. As a

consequence, good teachers can get bad evaluations, especially if they teach classes with a lot of bad students.

Consistent with this intuition, we also find that the evaluations of classes in which high-skill students are over-represented are more in line with the estimated quality of the teacher. Additionally, in order to provide evidence supporting the intuition that evaluations are based on students' realized utility, we collected data on the weather conditions observed on the exact days when students filled the questionnaires. Assuming that the weather affects utility and not teaching quality, the finding that the students' evaluations react to meteorological conditions lends support to our intuition.² Our results show that students evaluate professors more negatively on rainy and cold days.

There is a large literature that investigates the role of teacher quality and teacher incentives in improving educational outcomes, although most of the existing studies focus on primary and secondary schooling ([Figlio & Kenny, 2007](#); [Jacob & Lefgren, 2008](#); [Kane & Staiger, 2008](#); [Rivkin et al., 2005](#); [Rockoff, 2004](#); [Rockoff & Speroni, 2010](#); [Tyler, Taylor, Kane, & Wooten, 2010](#)). The availability of internationally standardized test scores facilitates the evaluation of teachers in primary and secondary schools ([Mullis, Martin, Robitaille, & Foy, 2009](#); [OECD, 2010](#)). The large degree of heterogeneity in subjects and syllabuses in universities makes it very difficult to design common tests that would allow to compare the performance of students exposed to different teachers, especially across subjects. At the same time, the large increase in college enrollment occurred in the past decades ([OECD, 2008](#)) calls for a specific focus on higher education.

Only very few papers investigate the role of students' evaluations in university and we improve on existing studies in various dimensions. First of all, the random allocation of students to teachers differentiates our approach from most other studies ([Beleche, Fairris, & Marks, 2012](#); [Johnson, 2003](#); [Krautmann & Sander, 1999](#); [Weinberg et al., 2009](#); [Yunker & Yunker, 2003](#)) that cannot purge their estimates from the potential bias due to the best students selecting the courses of the best professors. Correcting this bias is pivotal to producing reliable measures of teaching quality ([Rothstein, 2009, 2010](#)).

The only other study that exploits a setting where students are randomly allocated to teachers is [Carrell and West \(2010\)](#). This paper documents (as we do) a negative correlation between the students' evaluations of professors and harder measures of teaching quality. We improve on their analysis in two important dimensions. First, we provide additional empirical evidence consistent with an interpretation of such finding based on the idea that good professors require students to exert more effort and that students evaluate professors on the basis of their realized utility. Secondly, [Carrell and West \(2010\)](#) use data from a U.S. Air Force Academy, while our empirical application is

² One may actually think that also the mood of the professors, hence, their effectiveness in teaching is affected by the weather. However, students are asked to evaluate teachers' performance over the entire duration of the course and not exclusively on the day of the test. Moreover, it is a clear rule of the university to have students fill the questionnaires before the lecture, so that the teachers' performance on that specific day should not affect the evaluations.

Download English Version:

<https://daneshyari.com/en/article/6840906>

Download Persian Version:

<https://daneshyari.com/article/6840906>

[Daneshyari.com](https://daneshyari.com)