



UC3M Working Papers
Business
16-02
ISSN 2387-175X
July 2016

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249607

Teaching Quality and Academic Research

Rosa Rodríguez¹ and Gonzalo Rubio²

Abstract

This paper studies the relation between teaching quality and research productivity using a detailed database for students in higher education. In order to measure teaching quality, we employ a version of the value-added methodology which uses future performance of students to make inferences about the current teaching quality of their professors. We report a mild but positive and significant relation between publications on top journals and teaching quality. To all practical effects, this finding does not seem to be contaminated by the potential negotiating power of professors with high levels of top publications strategically choosing the best-performing groups. Additionally, we report evidence that the random assignment of students into class groups is reasonably satisfied. Finally, we argue that teaching effectiveness measures based on anonymous student evaluations are suspicious and debatable.

Keywords: Academic research; teaching quality; value-added; student questionnaire; standardized grade.

JEL classification: I20

¹ Department of Business Administration and Quantitative Methods, Universidad Carlos III, Calle Madrid 126, 28903 Getafe (Madrid), Spain; email: rosa.rodriguez@uc3m.es.

² Department of Economics and Business, Universidad CEU Cardenal Herrera, Plaza reyes Católicos 19, 03204 Elche (Alicante), Spain; email: gonzalo.rubio@uch.ceu.es.

Teaching Quality and Academic Research

(International Review of Economics Education, Vol. 23, 2016, 10-27)

Rosa Rodríguez

University Carlos III of Madrid

Gonzalo Rubio*

University CEU Cardenal Herrera

This version: May 30, 2016

Abstract

This paper studies the relation between teaching quality and research productivity using a detailed database for students in higher education. In order to measure teaching quality, we employ a version of the value-added methodology which uses future performance of students to make inferences about the current teaching quality of their professors. We report a mild but positive and significant relation between publications on top journals and teaching quality. To all practical effects, this finding does not seem to be contaminated by the potential negotiating power of professors with high levels of top publications strategically choosing the best-performing groups. Additionally, we report evidence that the random assignment of students into class groups is reasonably satisfied. Finally, we argue that teaching effectiveness measures based on anonymous student evaluations are suspicious and debatable.

Keywords: academic research; teaching quality; value-added; student questionnaire; standardized grade.

JEL classification: I20

We thank Victor Aguirregabiria, Raquel Carrasco, Belén Nieto, Alfonso Novales, José Penalva, Mikel Tapia, Robert Toutkoushian, and participants at the 10th Annual Meeting of Labor Economics at Complutense University in Madrid for helpful comments and suggestions that substantially improved the contents of the paper. We also thank the constructive and relevant comments of an anonymous referee. Rosa Rodríguez and Gonzalo Rubio acknowledge financial support from the Ministry of Economics and Competitiveness through grant ECO2015-67035-P, and Gonzalo Rubio from the Generalitat Valenciana grant PrometeoII/2013/015. We assume full responsibility for any remaining errors.

* Corresponding author: gonzalo.rubio@uch.ceu.es

1. Introduction

Universities around the world emphasize teaching and research as their main tasks. Consequently, the relation between academic research and teaching quality is one of the most fundamental issues of higher education.¹ This paper investigates this relation by employing not only student evaluations but also measures of teacher value-added effectiveness.

Although policy arguments assume that teaching is a key determinant of a student's academic formation and transition into the labour market, how to estimate teachers' contribution to student learning is much less clear. This fact is disturbing not only because empirical studies analyzing the relation between teaching quality and research may potentially be subject to well-founded causality critiques, but also because it makes the inclusion of teaching quality considerations into teachers' promotions and tenured positions extremely controversial.²

This paper applies value-added models of teaching effectiveness at the university level, with the final objective of analyzing the relation between teacher quality and academic research. As opposed to the value-added contemporaneous approach used in primary and secondary schools, this paper employs a version of the value-added methodology first proposed by Carrell and West (2010) which uses future performance of students to make inferences about the current teaching quality of their professors.

The issue of clarifying the relation between teaching quality and academic research is particularly relevant for countries where hiring companies do not tend to offer new positions for undergraduates based on the reputation of a particular university. In this context, where there seems to be a lack of market discrimination about the reputation of universities, policy administrators should be especially concerned with the budget resources assigned to public universities. This is particularly true for Continental Europe, where there is a current debate on whether university funding should be more strongly related to academic research contributions and less so to a school's number of students. The core of this debate relies on the degree to which teaching quality and academic research are interconnected. As discussed by Labini and Zinovyeva (2014), many universities worldwide have a complementary view of research and teaching. Teachers at the frontier of knowledge can then teach up-to-date material more effectively and their courses can better reflect the relevant issues to be taught. On the other hand, teaching and research

¹ See Mas-Colell (2011) for a discussion about the differences between universities in the United States and the European Union regarding the interaction between research and teaching as the two fundamental aspects to consider when thinking about the European space for higher education.

² Carrel and West (2010) discuss how students appear to reward teachers with good evaluations depending upon the contemporaneous grades they expect to get from the instructors. This finding seems to question the value and accuracy of student evaluations through anonymous questionnaires.

can also be considered substitutes, given the time and effort required for teaching excellence, and can even reflect different natural abilities. Ultimately, the relation between academic research and teaching quality is an empirical issue.

We study the relation between teacher quality and academic research for three degree programs, Business Administration, Economics, and Finance & Accounting, offered by the University Carlos III of Madrid, one of the best internationally known Spanish Universities, for two graduating classes from 2008 to 2013. We find a mild but positive and significant relation between academic research, as measured by the publications on top scientific journals, and teaching quality. It is important to point out that our findings do not seem to be explained by a non-random assignment of students to class groups, and to all practical effects they cannot seem to be attributed to the strategic actions of professors with top publications choosing the best-performing student groups for example taught in English, with high average entry scores or with a significant proportion of females. Additionally, we show that value-added-based measures of teaching quality seem to reflect different characteristics than the ones related to overall satisfaction, as captured by student questionnaires. In particular, and contrary to our previous finding, the degree of student satisfaction is negatively and significantly related to top publications, but it is strongly and positively related to teaching activities and to the grading methodology followed in the course. These results have serious implications for how student performance should be assessed and, of course, for how teacher quality should be measured.

We do recognize that we cannot generalize our results to the global higher education space. In this sense, our research has well-identified limitations. Although, our data are very rich and university officials have provided us with all the necessary details to pursue this research, the data cover only the School of Law, Business and Economics from a single Spanish university. The positive point is that the University Carlos III has a specific and well-defined system of incentives for both teaching and research that makes their data reasonable to use; however, it is very difficult to generalize our results to other areas of knowledge or to other Spanish universities. Nonetheless, a more ambitious goal of our study is a public call for the creation of comparable data sets across Spanish universities to allow researchers, students, and public official to discriminate between programs, schools, departments, and universities. At the very least, the top Spanish universities should develop systematic data sets that allow for well-defined questions and answers about the relative importance of teaching and research. They should understand that an investment in this direction is a key strategic tool in discriminating themselves from other universities while alleviating adverse selection problems.

From a more general perspective, studies of this sort, with public data sets that allow for discrimination in teacher quality, academic research productivity, employment conditions, recent graduates' salary levels, and so on, should be a fundamental part of the obligations of Spanish universities to society. Any society should be able to ask about the responsibilities of universities when enormous amounts of resources are invested in higher education. An obvious prerequisite of such studies is sufficient data to make the necessary comparisons and then penalize those who are not creating enough value in either teaching or research.

The paper is organized as follows. Section 2 briefly reviews previous literature on faculty research productivity and its relation with teaching quality. Section 3 describes our data regarding institutional details and discusses students and teacher characteristics related to research and teaching activities. Section 4 discusses the econometric methodology to obtain measures of teaching effectiveness. Section 5 explores the relation between academic research and teaching quality, employing a value-added methodology. Section 6 analyzes teacher evaluations from student questionnaires. Section 7 concludes the paper with summary and final remarks.

2. Previous Literature on Research Productivity and Teaching Quality

Most universities rely on summary statistics of student questionnaires to assess the contribution of teachers to student knowledge. Indeed, using the results from a 1999 United States (US) national survey for Economic departments, Becker and Watts (1999) show that student evaluations of teaching are, by far, the most widely way of assessing teaching quality. Not only that, but this procedure is often the only methodology employed to evaluate teaching effectiveness. Becker, Bosshardt and Watts (2011) show that, even using an updated national survey for 2011, the departmental evaluation practices have practically not changed except for the use of on-line questionnaires.

The idea of using student evaluation of teachers is, of course, to alleviate agency problems related to the fact that neither the quality of teachers nor their level of effort can be observed. Watchel (1998) argues that faculty opinions about the reliability of teachers' evaluations are extremely dispersed. If anything, the consensus seems to be rather critical about the use of anonymous questionnaires as a way of determining which university teachers to hire or to promote. Indeed, Carrel and West (2010), and Weinberg, Fleisher, and Hashimoto (2009) find that teachers who give higher grades receive better evaluations. More specifically, Carrel and West (2010) report that student evaluations are positively correlated with contemporaneous grades but negatively correlated with grades in subsequent courses.

Students seem to penalize professors who increase their deep knowledge improving their understanding to perform better in future courses. A similar finding is reported by Braga, Paccagnella and Pellizzari (2014 a) for Italian students at Bocconi University. In any case, even if these evaluation measures were effective, we would still confront the problem of causality. In other words, it is not clear at all that a positive correlation between measures of satisfaction and grades implies that teachers contribute positively to student learning.

Although these concerns about student evaluations are receiving increasing attention in the US university system, they do not seem to be a fundamental issue among university officials. Top universities are clearly discriminated at the market level, and general reputation arguments closely related to research activities play a key role in determining, at least indirectly, the effectiveness of teachers. Hence, it is not surprising that the academic literature identifying characteristics of teacher effectiveness has focuses on primary and secondary schools. This extensive literature analyzes the relative performance of schools and teachers comparing their value added to student learning. Value-added models generally measure the importance of teacher quality to educational production by standardized test scores. In these models, the outcomes scores are the sum of a teacher's effect, individual heterogeneity, and a transitory orthogonal error. Important papers in this literature are those of Goldhaber and Hansen (2010 a, 2010 b) who discuss the value-added methodology, showing a statistically significant relationship between teachers' value-added effectiveness measures and subsequent student achievement in their classes, Chetty, Friedman, and Rockoff (2014) who argue that primary students assigned to high-value-added teachers increase significantly their probability to attend college and earn higher salaries, and Rothstein (2010), who discusses a causal interpretation of the results. Rothstein argues that the estimates from value-added models can be interpreted causally only under unverifiable assumptions about the correlation between the assignment of students to teachers and other determinants of test scores. In practice, this implies that classroom assignments may not be exogenous, conditional on typical controls. Even if the best students do not self-select themselves into the classes of the potentially best teachers, Rothstein (2010) shows that estimators of teacher effectiveness can be substantially biased when selection is mostly driven by unobservable variables. Therefore, even if value-added estimates appear to be correlated with actual teacher effectiveness, they may not be unbiased.³

Despite numerous studies evaluating teacher effectiveness in primary and secondary schools and the well-understood statistical properties of their measures, evidence on the relation between academic

³ See Hanushek and Rivkin (2006, 2010) for a review and discussion of the value-added literature.

research and teaching quality at the university level is relatively scarce and the available results are inconclusive. Data are not available across universities worldwide, or even across different universities in a given country. The lack of publicly accessible data nationwide is striking. European public administrators and management of both public and private universities should be especially concerned with the lack of university interest in developing comparable data sets that might allow discrimination between universities, not only by research productivity but also by the connection between teaching effectiveness and academic research. Therefore, the results we have are specific to single universities, academic disciplines, or individual researchers. Data heterogeneity makes it difficult to generalize the evidence.

Hattie and Marsh (1996) summarize the evidence of previous studies in the US and conclude that there is a small but positive correlation between different measures of teaching quality and research, although the results vary greatly across the studies in their sample. More recently, Marsh and Hattie (2002) find a null relation between research and teaching, while Stack (2003) shows a positive and significant relation, as long as non-linearity is allowed in the econometric tests. Indirect evidence between research and teaching quality is also available from US universities. Using data from eight cohorts of first-year students at Northwestern University, Figlio, Schapiro, and Soter (2013) show that students seem to learn more from non-tenure track professors in their introductory courses. On the other hand, using institutional data from the College Board, Ehrenberg and Zhang (2005) show that the use of non-tenure track faculty does adversely affect graduation rates. Finally, using Canadian data and a value-added methodology to measure teaching effectiveness, Hoffmann and Oreopoulos (2009) conclude that there is not a strong correlation between research- and teaching-focused college instructors, since there are effective and non-effective teachers within each group. The authors suggest that, at the margin, instructors do not make a great difference in student achievement.

Regarding European data, Braga, Paccagnella and Pellizzari (2014 a) employ data from Bocconi University in Italy for students entering the 1998-1999 academic year in the Management, Economics, and Law and Management programs. Using a value-added approach, the authors conclude that professors who are more productive in research seem to be less efficient as teachers when research quality is measured by the H-index, a research contribution measure based on individual citations. The sign of the relation is reversed when the authors employ yearly individual citations but the evidence is not statistically significant. However, the impact of teaching effectiveness is relatively important. The average difference in subsequent performance between students assigned to the best teacher and to the

worst is approximately 5.6% of the average grade. García, Georgantzis, Martín, and Pérez (2012) employ the data of 604 university professors at the University Jaume I in Castellón, Spain from a variety of disciplines during 2002-2006. The authors find a significant non-linear positive effect between teacher quality and research performance when teacher effectiveness is measured by students' teaching evaluations.

Finally, the most exhaustive study relating academic research and teacher quality is that of Labini and Zinovyeva (2014) who employ a very rich data set from Italian universities. Their data cover around 26% of the total population of Italian university graduates from 2001 to 2004 and include outcomes in the labour market and in professional qualification exams. These data are combined with official statistical data about research quality and different measures of research output. The authors find that measures of student satisfaction with teaching are positively correlated with department-level indicators of academic research quality. Moreover and more importantly, the authors show that department-level variation of research productivity over time is positively correlated with cross-cohort variations in graduates' labour market outcomes.⁴ These outcomes are a very reasonable proxy for teaching quality and the results suggest the importance of systematically collecting these data from universities trying to distinguish themselves from potential competitors. Such data collection is a common approach followed by top business schools around the world, but it is much more difficult to find these statistics for either public or private European universities.

Given this rather disperse and conflicting evidence, our main contribution is that our results can separate alternative aspects of research. This implies that we have access to a very clean definition of what research means. Moreover, this measure is provided by the university itself which indicates that we do not have to construct ourselves measures of research. As we discuss next, our data distinguish two very different aspects of research. The first one over weights the number of years passed since the doctoral dissertation was presented while the second measure only takes into account the actual contributing papers in top journals. These precisely-defined measures of research production allow us to be clear on how deep learning improves students' performance. We do not know of any previous paper that incorporates alternative measures of research when studying its relation with teaching quality. And, even more importantly, this distinction of research tasks is available in a random allocation of students to courses as discussed in the next section.

⁴ Recently, Braga, Paccagnella, and Pellizzari (2014 b), employing data from Bocconi University, analyze the relation between academic and labor market returns of university professors. They find a positive but weak correlation between academic and labor market returns.

3. Data and Institutional Details

3.1 Institutional Data

The University Carlos III is a public Spanish university located in Madrid. It is a relatively recently founded university, by the standards of other, more traditional Spanish universities. The university started in the 1989-1990 academic year, offering only a bachelor's degree in Law. During the 2013-2014 academic year, the university had more than 18,000 students enrolled in a full variety of degrees in most areas of study. We have data from 4,613 students enrolled in the four-year degree programs of Business Administration, Economics, and Finance & Accounting. Table 1 reports the list of compulsory courses in each degree program. Our data include six academic years from 2008-2009 to 2013-14, corresponding to the first two complete graduating classes of students enrolled under the Bologna process or six full cohorts.⁵ The degree programs are taught at two different campuses located in the suburbs of Madrid. The main campus and larger campus are located in Getafe, but the students can also choose the smaller campus of Colmenarejo. Students have the opportunity of choosing a group taught in English. We obtained data on 66 compulsory courses and 541 professors in four different categories; external teachers, who are industry-related professionals with teaching responsibilities for courses for which they have a specialized level of expertise, tenure track professors, associate professors, and full professors.⁶

Table 2 shows the distribution of the teaching categories among the three degree programs for which we have data. Approximately 51% of the students are in Business Administration, 32% in Economics, and 17% in Finance & Accounting. Interestingly, given the usual standards of traditional Spanish universities, practically 61% of the students are taught by external teachers and 26% by tenure track professors. This implies that approximately only 13% of the students are taught by consolidated teachers with a guaranteed academic position at the university. This asymmetric distribution is maintained throughout all three degree programs. The teachers belong to five departments: Business, Economics,

⁵ The Bologna process started with the Declaration of Lisbon made by the European Union in 2000. At that point, the Executive European Council formally declared the will of the European Union to become "the most competitive knowledge-based economy in the world".

⁶ By courses, we mean subjects of study for which all students receive a grade, such as Mathematics for Economics I, Financial Economics, Econometrics, Financial Statement Analysis, Macroeconomics, International Trade, and so on.

Statistics, Economic History, and the Business Management Division (BMD) from the Mechanical Engineering department.⁷

Additional relevant information is the number of groups taught by each of the 541 professors available in our database. This variable changes significantly across professors. A “normal” professor teaches 10 groups during the full academic year (5 groups per semester). Each class lasts one hour and a half. This represents 8 hours of teaching per week during the year or 240 hours of teaching per year. However, a good research performance record decreases significantly the number of groups taught by teachers. Someone with the best research record teaches 6 groups per year. This means a 40% reduction due to outstanding research. Moreover, teachers with administrative responsibilities get reductions from 40% to 80% depending upon the position occupied by the teacher. Finally, women may have up to a 50% reduction in teaching responsibilities during the two years following maternity where the specific reduction depends on their research record.

The academic year at Carlos III consists of two terms (fall and spring) and four exam periods. The exams corresponding to the first term are taken in January and the exams for the second term are taken in June. Students who fail can take the exams again at the end of the academic year during the months of June (for those students who failed in January) and July (for those who failed in June). We employ data exclusively from grades obtained for exams taken for the first time, in both January and June of each academic year.

A total of 360, 190, and 110 students initially entered the degree programs in Business Administration, Economics, and Finance & Accounting respectively.⁸ Students of different degrees were allocated into different groups. Students in these groups have the same compulsory courses but usually different timetables and teachers. However, regardless of the specific group to which students are assigned, they are all taught the same material and all teachers follow the same syllabus. Additionally, all exams across groups contain the same questions. Students are allocated to a group based on their surname and are not allowed to change groups except for very strict reasons relating to an incompatible work schedule. All requests for changes have to be clearly justified, with corresponding documentation, to the vice-dean of the degree program. This procedure indirectly generates a random allocation process of

⁷ Given the different views on how to interpret the research productivity of teachers in the Law department, we exclude both Law courses taught in the Business, Economics, and Finance & Accounting degree programs and Law professors from our sample.

⁸ These numbers are exact for the 2013-2014 academic year. During the different years covered by our sample period, this number varied slightly during the first years and we should also take into account new students from other degree programs or universities.

students to different teachers. This is important because it facilitates the interpretation of our empirical results. When analyzing the relation between measures of teaching effectiveness and research quality, it becomes extremely complex to distinguish between correlation and causality effects. The precise way in which the University Carlos III assigns students across groups and professors, based on their last name, represents a great advantage of our data. It guarantees, to a certain degree, that the assignment of students to groups is random. This is true for the first year of their degrees but it is also true for the rest of the years in the different degree programs. As already pointed out, we measure teaching effectiveness at time t with performance in exams in time periods $t + 1, t + 2, t + 3 \dots$. This procedure includes teachers in all years of study which may incorporate undesirable cumulative learning effects. However, the impact of cumulative learning, given our database, is limited. This is because the allocation of teachers in year 2 is random for students who were taught in year 1 by a particular professor. A given professor does not necessarily teach the same students because they are generally distributed among groups with an alphabetic order, and they do not have to be assigned to students with last names starting with the same letter as previous years. The same thing applies for students of years 3 and 4.

Table 3 shows descriptive statistics that summarize the distributions of compulsory courses and their groups across terms and degree programs. During the initial years of the sample, the average class sizes of Economics and Business Administration are slightly higher than in Finance & Accounting. This tendency changes from the 2011 academic year onwards. The average class size in Finance & Accounting becomes slightly higher than in Economics and the Business Administration degree starts having a larger class size except for the second term in 2013 where the largest class size is in the Finance & Accounting degree. Overall, however, we may argue that the average class size is similar across all degree programs. This is an important characteristic, given the evidence of Pinto and Vera (2011), who employ data from student enrolled in the first year of compulsory courses for the Business degree at Carlos III from 2000-2001 to 2005-2006. The authors find that class size negatively affects only medium-ability students. Overall, they show that the medium-ability students seem to be the most sensitive to class and peer effects relative to high- or low-ability students.

3.2 Student, Teaching, and Research Performance Data

Our data cover the entire academic history of students in three degree programs for the two available graduating classes. We have all their grades for all compulsory courses taken in their four years at the university, as well as their overall personal and demographic characteristics. Thus, we have their gender,

age, average entry test grade, place of residence before coming to the university, and whether they came from a regular high school or have a professional formation degree.⁹ We also know the group each student is assigned to for each of the courses taken. Table 4 reports descriptive statistics for the students in our sample for all three degree programs. Economics is the only program with slightly more male than female students, while the highest average entry score is for students entering in Business program.¹⁰ Finance & Accounting is the least successful degree program in attracting students from outside Madrid.¹¹ The percentage of students entering from the standard nation-wide examination is similar across all three degree programs. Unfortunately, we have no data on the initial labor market conditions for our sample students. The availability of such data would facilitate the analysis of teaching effectiveness, as in Labini and Zinovyeva (2014) and Braga, Paccagnella, and Pellizzari (2014 b).

We also have data regarding students' teacher evaluations. Before ending the term and before receiving their final grade, students are asked to evaluate their teachers through the use of standard questionnaires. These evaluations are anonymous but we were able to identify the name of the course, the group identifier, and the name of the professor. We obtained an overall evaluation of the teacher, as well as the opinions about the course evaluation methodology employed. Table 5 shows the average results for various departments and years. Each item in the questionnaire asked the students' rating on a scale from one (very negative) to five (very positive). Overall satisfaction, on average, does not seem to follow any well-defined pattern throughout the years or the departments. If anything, it can be argued that for the first year of studies, teachers from the Business department, the BMD division of the Mechanical Engineering department, and the Economic History department tend to obtain slightly better evaluations than teachers from Economics, Statistics. This pattern is reversed for the second, third and fourth year of studies where the Economics department obtains the highest average evaluation from students. Generally speaking, the Statistics department tends to get the lowest average evaluation with the important exception of the second year of studies in 2011. In any case, the differences across years or departments do not seem to be either high or consistent enough to infer overall clear-cut conclusions.

⁹ The average entry score is a weighted average of high school degrees and grades obtained in the national university admission test, an exam that has been traditionally the same for all Spanish students. In the years covered by this research, students had the opportunity to take two additional exams to improve their entry grades. We will come back to this issue later in the paper.

¹⁰ The average entry score is calculated over a maximum of 14 points where 10 points come from the same nation-wide exam and 4 points from the two additional exams that students may take.

¹¹ We must point out that this was the first time that a degree in Finance & Accounting was offered by any Spanish university. The fact that this degree was relatively unknown at that time may explain this characteristic.

Our sample contains data on individual teachers' age, gender, category, and whether they have a PhD or not as well as official university indices on research productivity and teaching activities. Every year, the university runs an internal official competition that ranks each professor with a doctoral degree in terms of research activities, top publications, and teaching activities. The university administrators try to incentive the necessary effort and dedication to both research and teaching. An additional economic benefit is obtained, depending upon the position reached in the ranking. Research activity includes research periods approved by the Spanish government, both the absolute number of periods and the relative number of periods, the number of supervised doctoral dissertations, and competitive research projects participated in as a principal researcher.¹² Top publication activity counts the number of publications in journals listed in the *Journal of Citation Report* (JCR), where the points obtained depend upon the journal's JCR quartile with the first, second, third, and fourth quartiles receiving eight, four, two and one point, respectively. Note that for both, research and publication activities, all teachers holding a PhD can access to these rankings. Finally, teaching activity is composed of overall satisfaction obtained from the students' evaluations, time employed to correct exams and give the final grades of courses, activities directed toward innovative teaching projects, the elaboration of teaching material, and the coordination of courses within the department. As before, all teachers with a PhD have access to this ranking. Table 6 shows descriptive statistics and the range of variables employed in this paper for degree programs, students and teachers.

4. Estimation of Teacher Quality

We first estimate a measure of teacher effectiveness or quality by obtaining the conditional mean of future grades in compulsory courses, given that a course c_0 has been taught by a particular professor. The traditional value-added approach employed in primary and secondary schools use contemporaneous grades. At the university level, however, it is important to use future performance to obtain current teaching quality measures. This is particularly relevant given the concerns pointed out before about the inferences from questionnaires where student evaluations may be endogenous relative to student grades.

¹² Every university teacher in Spain has the opportunity to be evaluated every six years in terms of his or her research contributions. A non-anonymous committee evaluates the five most relevant publications of every teacher in a six-year period. Teachers from both public and private universities can compete on a similar basis. The number of recognized periods is simply the number of the six-year periods every professor has, while the relative period is the ratio of recognized periods to the maximum possible number of periods the professor may have obtained in his or her academic career.

Moreover, as discussed by Carrel and West (2010), standardized tests are not given at the university level, and grades depend on heterogeneous assignments and tests.¹³

As previously discussed, we have data on three different degree programs Business Administration, Economics, and Financial & Accounting, denoted by $d = 1, 2, 3$. Each student $i = 1, \dots, I_d$, in one of the three degree programs, attends a given sequence of compulsory courses $c = 1, \dots, C_d$, where C_d is the total number of compulsory courses in degree d . In each of these courses, every student is allocated to a class or group $g = 1, \dots, G$, where G_c denotes the total number of groups for all courses c in each program d , according to the first letter of the student's surname. For a given group g taught by a particular (identified) professor, we run the following panel regression, where the dimensions of the panel are the individuals and the future courses they take:

$$Y_{ic}^g = \alpha^g + X_i^g \beta + \varepsilon_{ic}^g, \quad (1)$$

where Y_{ic}^g is the grade obtained by student i in course c , where c is a compulsory course different from course c_0 , which is taken by the student in any future quarter following the specific quarter in which course c_0 was taken. The grades across courses, Y_{ic}^g , are standardized at the group level to control for potential differences in grade distribution. The explanatory (control) variables, given by the components of X_i^g , are student characteristics that include entry test score, a gender dummy, a dummy for the student's geographic origin (outside Madrid or not), age, and temporal dummies. The estimated alphas can be interpreted as the conditional means of the future grades of students in group g . In other words, high (low) values of $\hat{\alpha}^g$ suggest that, on average, students attending course c_0 in group g (or, similarly, taking the course c_0 with professor $j = 1, \dots, G$) obtained a better (worse) grade in subsequent courses than students attending course c_0 in a different group (with a different professor).¹⁴ Assuming that the students are effectively randomly allocated across groups, we can interpret the estimated intercepts of the previous panel regression as exogenous group fixed effects. We estimate the panel by the standard random effects procedure with the Swamy-Arora (1972) estimator of the variance of the individual and

¹³ See also the analysis of Braga, Paccagnella, and Pellizzari (2014 a, b).

¹⁴ We eliminate the few groups taught by two or more different professors.

the purely random components. We estimate 1,955 fixed effects and therefore run 1,955 panel regressions with random effects.

Estimates of the intercepts (alphas) can be contaminated by the characteristics of the specific group a professor is teaching. Therefore, we must purge the previously estimated intercepts/alphas of group characteristics that may influence the future performance of students but that cannot be attributable to teachers. We first point out that our next analysis is performed by jointly conditionally on degree, location of the program, and language in which the program is taught. The analysis also controls for the average entry score of the students in that group, the percentage of women in the group, the proportion of students entering the university in ways other than the usual selection process of Spanish universities, the group's size, the proportion of students from outside Madrid, and the average age of a group's students. These characteristics can all potentially impact the estimated alphas over and beyond the professor's abilities. For this reason, jointly conditional on degree, location, and language, we next perform the following heteroskedasticity-robust ordinary least squares (OLS) regression of the estimated alphas on the group characteristics:

$$\hat{\alpha}_{adj}^g = X^g \beta + u^g, \quad (2)$$

where $\hat{\alpha}_{adj}^g$ is the 1,955-dimensional vector of the estimated alphas in equation (1) adjusted by the inverse of the standard errors of the corresponding alpha coefficients, and X^g is the vector of group characteristics included in the list mentioned above. We divide the estimated alphas by their standard errors of regression (1) to take into account the differences in the precision with which the alphas are estimated.

After these controls, the remaining factor explaining the cross-sectional variation of the estimated alphas should be associated with teaching effectiveness. This approach suggests that the residuals from regression (2) can be interpreted as the unobservable teaching quality. The residuals denoted by u^g are therefore the measure of teaching quality employed in this paper, where the conditional means of future grades are purged from student and group characteristics. They measure teachers' ability to influence the future performance of their students.¹⁵

¹⁵ As in all previous literature, the causal interpretation of these residuals requires making a number of important assumptions. These assumptions go from the functional form of the regression to the existence of a mechanism of random assignment of students and professors to different groups. As pointed out in Section 3, we try to take advantage of the surname procedure employed by the university to assign

Table 7 presents descriptive statistics for the residuals of regression (2). These statistics are given for the overall data in our sample and by degree program, category of professor, year of study and department. A common way of measuring teaching effectiveness in the value-added literature consists of using as a metric the standard deviation of the estimates of teacher effects from our models. These standard deviations are a measure of how much teachers vary in their effects on performance or, in our case, future performance. Our overall results suggest that one standard deviation increase in teacher quality results in a 1.67 standard deviation increase in student achievement. The empirical results seem to be consistent across degrees, teacher categories, years of study, and department with the exception of full professors who show a lower standard deviation although the number of professors in the sample is much smaller than for the rest of teacher categories.

5. The Relation between Academic Research and Teaching Effectiveness

We now analyze the main research question of this paper. To study the relation between teacher quality and academic research productivity, we run the following heteroskedasticity-robust OLS regression (jointly conditional, as before, on degree, location, and language) of the 1,955 estimated adjusted alphas on the controls $X^{g,t}$:

$$\hat{\alpha}_{adj}^g = X^{g,t} \beta + v^g \quad (3)$$

where $X^{g,t}$ now contains not only the previously discussed group characteristics but also the teacher characteristics age, research, top publications, teaching activities, and a dummy variable for category of professors.

The first two columns of Tables 8 show evidence from regression (2) with the full sample of 1,955 where the difference between both columns reflects the use of term dummies in the second column.¹⁶ An important issue is concerning this estimation is that there are repeated individuals within the sample. This is clear since we have 1,955 groups but not 1,955 academics. This is why the p -values reported in Table 8 are based on standard errors clustered at teacher level. Otherwise, the standard errors would probably be too low, increasing the significance of the results. The empirical evidence contained in the second column

students into classes. Later in the paper, we actually test both the random allocation of students and professors.

¹⁶ The results reported in Tables 8 and 9 are from regressions without intercept. The qualitative final results regarding teaching quality and the relation between academic research and teaching effectiveness are very similar when we employ the results from the same regressions with an intercept. The results are available from the authors upon request.

of Table 8 shows that the average entry scores, the number of students from outside Madrid, and the average age of the group are significantly correlated with our measure of teaching quality. On the other hand, the proportion of females, the students from non-academic high schools, and the size of each group does not seem to be correlated with conditional future grades. In particular, a key factor seems to be the average entry score which is negative and statistically significantly correlated with teaching quality. Conditional futures grades tend to be lower the higher the average entry score. This is a surprising but interesting result. The entry requirements to public Spanish universities changed during the sample period employed in this research relative to previously requirements. The students in our sample have the opportunity to take two additional exams over and above the general compulsory exam. This suggests that students without the required initial grade have the opportunity to choose freely two additional exams to improve their entry grade. The consequences of this new policy have not been discussed anywhere, but the results of Table 8 calls for a serious investigation of this issue. At the same time, the number of students from outside Madrid is positively correlated with conditional future grades which make sense since these are probably good students willing to make the necessary effort to study in a competitive university away from home. Finally, the lower the average age of the group the better future conditional grades are. This result also makes sense since, among the older students in each class, there seem to be a higher proportion of students simultaneously working and studying.

The third and fourth columns of Table 8 report the relation between academic research and teaching quality. Therefore, we now analyze how teacher characteristics associate with research and teaching activities influence the future performance of the students in our sample. This is the key research question of this paper. Before discussing the empirical results, we must address a very relevant research issue in the educational literature that may affect the correct specification of our regression model. Staiger and Rockoff (2010) employ data from the two largest school districts in the U.S. Among other findings, they report that teachers' effectiveness rises rapidly in the first year or two of their academic careers but then flatten out. As they point out, this finding has serious policy implications for the hiring and retention process of good teachers. This suggests that whilst teaching quality improves with experience, the learning trend is not linear. The main consequence is that it would be inappropriate to employ age experience as one of the control variables. By imposing a linear experience trend, the relation between teaching quality and publications may be biased because of the specific relation between age and publications. In Table 8, and in order to recognize the non-linear trend, we include not only the teachers'

age but also the square of age as a proxy for teaching experience. This recognizes the increasing marginal benefits from teaching experience but in a non-linear fashion.

Once we control for student and group characteristics, teaching quality, measured by the conditional mean of future grades, seems to be mildly but positive and significantly related to top publications. In the third and fourth columns of Table 8, the estimated coefficients associated with highly productive research are positive and with *t*-statistics of 1.83 and 1.96, depending upon the specification employed in the analysis. Interestingly, research activities do not explain teacher effectiveness, and a similar finding is obtained for teaching activities. The different correlations with respect to teaching quality obtained between top publications and research activities represent a potentially important result. To the best of our knowledge, this is the first empirical evidence regarding this issue. It is not the officially recognized research activities what is important for teaching, but the quality of research. On the other hand, it is important to note that the teaching activity measure includes the overall satisfaction obtained from the students' evaluations in addition to the time employed to correct exams, and other activities related to innovative teaching projects. Given this result, it may be useful that university authorities may revise the criteria employed in constructing the teaching activity measure. We will further discuss this issue later in the paper.

It is interesting to point out the lack of statistical significance of research activities. It may be surprising that the positive relation found with respect to top publications does not generate a positive relation with research activities. It must be recalled that research activities do not take into account publications. The three activities used to measure research activities at the University Carlos III, namely, doctoral thesis supervisions, to be a research principal in public research projects and to have officially recognized six-year research periods, imply a relatively long working period as a PhD university professor. In fact, the low average for research activities reported in Table 6 reflects precisely this point. This suggests that we may have teachers with top publications that obtain a very low number under research activities. Therefore, this item may reflect more the number of years working for the university than really the contribution to highly-recognized research productivity which is what it seems to be relevant for increasing teaching quality.

An important characteristic of the model employed to obtain teaching quality measures is that the independent variables have a linear relationship with the dependent measure. When analyzing the relation between teaching quality and top publications, we

have used the full sample with 1,955 groups. However, it turns out that only 760 groups have teachers who are candidates for research activities, top publications, and teaching activities. External teachers are not evaluated by university officials.¹⁷ This implies that roughly 61% of the teachers get zero research or teaching activity measures. This skewed distribution suggests a non-linear specification in the regression given by expression (3). In order to check whether the results reported in the third and fourth columns of Table 8 are badly specified, we also run regression (3) adding a squared term to top publications. The results are contained in the fifth and sixth columns of Table 8. The results confirm the importance of top publications. In fact, the positive correlation reported above between teaching quality and top publications is maintained, although the weak statistical significance remains with t -statistics of 1.79 and 2.06 depending upon the specification used in the regressions.¹⁸

Given the potential importance of teaching experience, we should be confident about the way in which we incorporate a non-linear experience trend in the analysis. In order to check the results reported in Table 8, and following Rockoff (2004), we propose an alternative proxy for non-linear experience which given by,

$$Exp_{jt} = \begin{cases} Age_{jt} & \text{if } Age_{jt} \leq \bar{Age} \\ \bar{Age} & \text{if } Age_{jt} > \bar{Age} \end{cases} \quad (4)$$

¹⁷ To be an external teacher, it is necessary to work outside the university system. External teachers are professionals working in the industrial or financial private or public sectors. The idea is to complement the academic education provided by internals. They may have a Ph.D. and they may even have academic publications. However, they are not evaluated by university officials and, therefore, we do not have data on their research activities and publications.

¹⁸ As a robustness check, we repeat the analysis estimating the relation between teaching quality and academic research using only 760 groups taught by individuals with research evaluations. The results are very similar in all cases. If any, the positive relation between teaching quality and top publications is now even stronger. The t -statistics are 1.97, 1.63, 2.19, and 1.80 for the last four columns of Table 8, respectively.

where Age_{jt} is the age of teacher j at time t , and \bar{Age} is the average age across teachers in the sample.

The results, reported in Table 9, are practically identical.¹⁹ There is a mild but positive and significant relation between top publications and teaching quality. However, none relation is found between the overall measures of research activities and teaching effectiveness. The main finding of the paper remains independently of how we control for teaching experience.²⁰

We recognize that our positive and significant relation between teaching quality and top publications may be contaminated if the allocation of students or teachers to groups is not random. Unless, we are able to show that these assignments are random, it is not obvious how we should interpret these coefficients. Ideally, we would like to make a causal interpretation from our regression estimates but the assumptions needed to infer causality are unlikely to be satisfied. Therefore, this is a key point to resolve before we can interpret our evidence as a causality relation, rather than just as a simple correlation arisen due to the possibility of students choosing the best professors or to the negotiating power of full professors strategically choosing the best-performing groups.

To formally test whether the allocation of students or teachers to alternative groups is random, we follow the procedure employed by Braga, Paccagnella, and Pellizzari (2014 a, b). We first analyze the random assignment of students to groups. The idea is to check whether students' observable characteristics are randomly balanced across classes. The statistics reported in Table 10 are derived from probit (Non Academic High schools, Gender and Outside Madrid characteristics) or OLS (Age and Entry test score characteristics) regressions of this observable student's characteristic on class dummies for each course in each degree program and academic year that we consider. The reported p -values refer to test of the null hypothesis that the coefficients on all the class dummies in each model are all jointly equal to zero. The test statistics are either a Chi-square or an F with varying parameters depending on the model, and the results are reported for the three degree programs individually and for the full sample in the last column of Table 10. For example, in Economics we have 8 courses during the first year times 6 academic years for a total of 48 courses, 8 courses during the second year times 5 academic years, and 5 courses during the third year times 4 academic years. This makes a total of 108 given in the last row of Table 10 under the Economics label. The same reasoning applies for Finance & Accounting and Business Administration. Adding 210, 108, and 117 makes a total of 426 for our full sample. Overall, we conclude that the

¹⁹ The results reported in Table 9 also employ standard errors clustered at teacher level.

²⁰ A cubic non-linear adjustment generates similar results with a slightly higher t -statistics. The results are again very similar when we employ only 760 groups taught by teachers with research evaluations.

students' available characteristics are well balanced across groups. The most problematic case is the entry test score in which we have 73 cases out of 426 for which we reject the equality of means across classes. This implies that in 17.1% of the cases we reject the null hypothesis of equal means. In fact, in Finance & Accounting and Business Administration the percentages are 14.9% and 6.8% respectively, which suggests that the entry test score in Economics with 32.4% of rejection is the only case in which the random allocation of students seems to be debatable. In any case, it is important to remember that the entry score is a bad predictor of future conditional grades.

Although we may argue that the evidence suggest that students are randomly allocated to classes, we also have to investigate whether teacher are randomly assigned across groups. We next perform a seemingly unrelated simultaneous equations system where we regress teachers' observable characteristics on the groups' observable characteristics. In particular, we estimate five seemingly unrelated simultaneous equations, where the observations are the 1,955 available groups for compulsory courses. The dependent variables are the five teacher characteristics, namely, research activities, top publications, teaching activities, age, and the category of the professor. The explanatory variables are the group characteristics given by the location of the program, language, average entry score, average age of groups, proportion of females and students from outside Madrid, class size, and years of study.

Table 11 reports the empirical results of random allocation tests for teachers. The reported F -statistic of Panel A tests the null hypothesis that the coefficients of each group characteristic are all jointly equal to zero in each equation of the system. The last row of Panel A tests the null hypothesis that the coefficients of all independent variables are jointly zero in all equations of the system.

Overall, the results show that we can reject the random allocation of teachers among groups. For the location of the program, language, average age, students from outside Madrid, class size, and the year dummies, the correlation with teachers' characteristics is statistically significant. However, this finding does not necessarily invalidate our conclusion regarding the positive relation between teaching quality and top publications. Our concern is that full professors strategically choose the best-performing groups. In other words, the key point, to validate the positive relation between teaching quality and top publications, is to analyze the statistical significance of the groups' characteristics with respect to the teachers' characteristics affecting the positive relation between publication levels and teaching quality.

Panel B of Table 11 shows the t -statistics from the seemingly unrelated individual equations of the simultaneous system of equations. We may think that teachers with high publication measures choose groups taught in English, which are usually believed to be composed of high-performing students, or

groups with particularly high average entry scores, or classes offered during the last years of the degree programs. The results reported in Panel B show that top publications are not statistically related to language, entry score, or positively related to the proportion of students from outside Madrid, and to the third year of the program. It is true however, that teachers with high level of publications seem to be assigned more often than expected to groups with younger students. This is a disturbing finding because groups with younger students are reported to have higher future conditional grades. However, it is difficult to imagine top researchers looking for the age of the students assigned to each group before deciding which group to teach. Finally, teachers with high a publication rank tend to teach in groups located in the main campus of the university.

To conclude, we show that the positive and statistically significant relation between top publication and teaching quality reported in Tables 8 and 9 does not seem to be contaminated by the strategic decisions of professors with high levels of productive research that may have enough negotiating power choosing ex-ante the most convenient or the best-performing groups of students. The only disturbing result is associated with the average age of the groups which is hard to imagine as a strategic decision variable among top researchers.

6. Teacher Quality and Student Questionnaire Evaluations

This section investigates the determinants of the overall student satisfaction from the evaluations that all professors receive from their students. As discussed in Section 4, the standard questionnaire used by the university evaluates teachers on a scale between one and five. From these questionnaires, we collect data on overall student satisfaction with each teacher and on the agreement of the way teachers evaluate their students.

Table 12 shows the results from a fixed effects panel data model where the dependent variable is the overall satisfaction of the students from 1,955 groups with instructors that may have their research activities measured by university officials. We also use the same data to estimate a generalized least squares random effects panel data model. It turns out that the Hausman (1978) test rejects the null correlation between individual effects and regressors, so we report the results from the fixed effect specification and, as before, we jointly conditional on degree, location, and language. The first column of Table 12 reports the results of directly using the standardized grades received by the students as one of the independent variables. The second column employs the residuals of the standardized grades as one of the independent variable, where these residuals are obtained from a previous regression of the

standardized grades on the degree program, program location, evaluation year, group to which the student is assigned, entry score, gender, student age and origin and whether the teaching language was English or not.

Both columns show a positive but non-significant relation between the overall satisfaction and the grades obtained by the students. It may be recalled that the questionnaires are prepared before the students received their grades. The variable estimated with the highest precision is the method employed to grade the students. This variable is strongly and positively correlated with the overall level of satisfaction. This is an important result because it generates doubts about the validity of students' questionnaires. Additionally, the younger the professor the less satisfied the students are. The most relevant result, given the evidence reported previously, is that both research activities and top publications are negatively and significantly correlated with the level of satisfaction. It seems that students do not internalize the impact on deep learning that obtain from teachers with top publications. Moreover, and contrary again to the previous results, there is positive and statistically significant correlation between the level of satisfaction and teaching activities. It seems that the influence of these activities have an immediate impact on the perception the students have about their professors but, as discussed before, these activities do not seem to have a persistent impact on future knowledge. This finding suggests that students perceive very differently what they are really learning relative to the apparent satisfaction they obtained from the professors. It seems that how the students are graded and teaching activities are key points reflected in their teacher evaluations which may distort the impact of other more important variables from the point of view of their learning process over time.

7. Conclusions

The relation between teaching quality and academic research is a very complex issue. European universities, especially in public-based university systems such as in Spain, should be involved in developing detailed databases that allow comparable analyses across universities. The results reported in this paper suggest that teaching quality measured by student questionnaires may be contaminated by the procedures employed to grade them and the associated teaching activities that have a questionable value on the learning process over time.

We propose measures of teaching quality based on the value-added methodology traditionally employed in non-university studies. However, contrary to the measures used in primary and secondary schools, we use conditional means of future grades received by students as the basic measure to evaluate

teaching quality. Then, we purge these conditional means by the individual characteristics of students and the characteristics of the groups to which they are assigned. Whatever is left should be attributable to teacher quality. Using data from the University Carlos III of Madrid from 2008 to 2013, we find that teaching quality is weakly but positively and significantly related to top publications, although we do not find a significant relation with respect to either research or teaching activities. To all practical effects, this positive relation does not seem to be contaminated by the potential strategic negotiating power of full professors choosing the best-performing groups or by the assignment of students to groups. The knowledge and experience obtained in the subject being taught through highly productive research published in top journals contributes positively to the learning process of students at the University Carlos III of Madrid.

References

- Becker, W., M. Watts (1999), How Departments of Economics Should Evaluate Teaching, *American Economic Review (Papers & Proceedings)* 89, 344-349.
- Becker, W., W. Bosshardt, and M. Watts (2011), Revisiting How Departments of Economics Evaluate Teaching, Working Paper, Purdue University.
- Braga, M., M. Paccagnella, and M. Pellizzari (2014 a), Evaluating Students Evaluations of Professors, *Economics of Education Review* 41, 71-88.
- Braga, M., M. Paccagnella, and M. Pellizzari (2014 b), The Academic and Labor Market Returns of University Professors, Working Paper 981, Banca d'Italia, Eurosisistema.
- Carrell, S., and J. West (2010), Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors, *Journal of Political Economy* 118, 409-432.
- Chetty, R., J. Friedman, and J. Rockoff (2014), Measuring the Impacts of Teachers II: Teacher Value-Added Estimates, *American Economic Review* 104, 2593-2632.
- Ehrenberg, R. and L. Zhang (2005), Do Tenured and Tenure-Track Faculty Matter?, *Journal of Human Resources* 40, 647-659.
- Figlio, D., M. Schapiro, and K. Soter (2013), Are Tenure Track Professors Better Teachers? National Bureau of Economic Research, NBER Working Paper 19406.
- García, A., N. Georgantzis, J. Martín, and T. Pérez (2011), (How) Do Research and Administrative Duties Affect University Professors' Teaching?, <http://ssrn.com/abstract=2133898>.
- Goldhaber, D., and M. Hansen (2010 a), Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions, Working Paper 31, CALDER, The Urban Institute.
- Goldhaber, D., and M. Hansen (2010 b), Using Performance on the Job to Inform Teacher Tenure Decisions, *American Economic Review (Papers & Proceedings)* 100, 250-255.
- Hanushek, E., and S. Rivkin (2006), Teacher Quality, in *Handbook of the Economics of Education*, ed. By A. Hanushek, and F. Welch, North Holland, Vol. 1, 1050-1078.

- Hanushek, E., and S. Rivkin (2010), Generalizations about Using Value-Added Measures of Teacher Quality, *American Economic Review (Papers & Proceedings)* 100, 267-271.
- Hattie, J., and H. Marsh (1996), The Relationship between Research and Teaching: A Meta-Analysis, *Review of Educational Research* 66, 507-542.
- Hausman, J. (1978), Specification Tests in Econometrics, *Econometrica* 46, 1251-1271.
- Hoffmann, F., and P. Oreopoulos (2009), Professor Qualities and Student Achievement, *Review of Economic and Statistics* 91, 83-92.
- Labini, M., and N. Zinovyeva (2014), The Relationship between Academic Research and Teaching Quality, Working Paper, IMT Lucca Institute for Advanced Studies and FEDEA.
- Mas-Colell, A. (2011), The European Space of Higher Education: Incentive and Governance Issues, Fifth Angelo Costa Lecture, Pompeu Fabra University.
- Marsh, H., and J. Hattie (2002), The Relationship between Research Productivity and Teaching Effectiveness: Complementary, Antagonistic or Independent Constructs? *Journal of Higher Education* 73, 603-641.
- Pinto, M., and M. Vera (2011), Peer Effects and Class Size in College, Working Paper, University Carlos III and University College London and Institute for Fiscal Studies.
- Rockoff, J. (2004), The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data, *American Economic Review* 94, 247-252.
- Rothstein, J. (2010), Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement, *Quarterly Journal of Economics* 125, 175-214.
- Stack, S. (2003), Research Productivity and Student Evaluation of Teaching in Social Science Classes: A Research Note, *Research in Higher Education* 44, 539-556.
- Staiger, D. and J. Rockoff (2010), Searching for Effective Teachers with Imperfect Information, *Journal of Economic Perspectives* 24, 97-118.

Swamy, P. and S. Arora (1972), The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models, *Econometrica* 40, 261-275.

Wachtel, H. (1998), Student Evaluation of College Teaching Effectiveness: A Brief Review, *Assessment & Evaluation in Higher Education* 23:2, 191-212.

Weinberg, B., B. Fleisher, and M. Hashimoto (2009), Evaluating Teaching in Higher Education, *Journal of Economic Education* 40, 227-261.

Table 1
Compulsory courses, by degree program

Finance & Accounting	Economics	Business Administration
Year 1 – Term 1		
Economic History		
Introduction to Business Administration	Introduction to Business Administration	Introduction to Business Administration
Principles of Economics	Principles of Economics	Principles of Economics
Mathematics for Economics	Mathematics for Economics I	Mathematics for Economics I
Year 1 – Term 2		
Financial Mathematics	Mathematics for Economics II	Mathematics for Economics II
Microeconomics	Microeconomics	Microeconomics
Introduction of Accounting Statistics I	Accounting I Economic History Statistics I	Introduction of Accounting Economic History Statistics I
Year 2 – Term 1		
Financial Economics	Macroeconomics	Macroeconomics
Statistics II	Statistics II	Statistics II
Organization and Management	Game Theory Microeconomics Theory	Game Theory Management Accounting
Year 2 – Term 2		
Econometrics I	Econometrics	Econometrics
Financial Accounting I	Financial Economics	Financial Economics
Fixed Income and Derivatives	Dynamic Macroeconomics	Organizational Behavior
Financial Management	Industrial Organization	
Year 3– Term 1		
Financial Accounting II	Econometric Techniques	Marketing
Econometrics II	Public Economics	Organizational Economics
Cost Management	Markets and Environment	
Financial Institution System		
Year 3 – Term 2		
Financial Statement Analysis	Applied Economics	Marketing Management
Financial Risk Management	International Trade	Financial Management
Management Control		
Year 4 – Term 1		
Auditing Financial Statements		Strategic Management
Corporate Management and Social Responsibility		Operations Management

This table reports the list of courses used in this paper to analyze the relation between teaching quality and academic research. The courses are listed by academic degree program, semester and year in which these courses are taught.

Table 2
Percentage of students taught, by professor category, and degree program

Degree Program	External	Tenure-track	Associate	Professor	%Students
Finance & Accounting	9.01 (14.88)	3.95 (15.17)	3.71 (36.21)	0.50 (16.02)	17.17
Economics	19.59 (32.34)	9.39 (36.01)	1.41 (13.76)	1.29 (41.22)	31.67
Business Administration	31.97 (52.79)	12.72 (48.81)	5.12 (50.03)	1.34 (42.77)	51.16
Total	60.57	26.07	10.23	3.13	100.00

The last column of this table shows the percentage of students chosen a given degree program, while the last row contains the percentage of each category of professors teaching in these three academic programs. The first row, for a given program, contains the percentage of students taught in each degree program by a given category of professors. For example, out of the 17.17% of students that choose Finance & Accounting, approximately 9.0 are external teachers. In parentheses, we report the percentages of teachers' categories in the different degree programs. For example, out of the 60.57% of external teachers, approximately 53 teach in Business Administration, 32 in Economics and 15 in Finance & Accounting.

Table 3
Descriptive statistics of degree programs

	Finance & Accounting			Economics			Business Administration		
	# Courses	# Groups	Class Size (Average)	# Courses	# Groups	Class Size (Average)	# Courses	# Groups	Class Size (Average)
2008									
1 st Term	4	20	28	3	22	30	3	43	27
2 nd Term	4	20	27	5	40	31	5	73	26
2009									
1 st Term	7	36	23	7	64	28	7	102	28
2 nd Term	8	42	20	9	85	26	8	114	27
2010									
1 st Term	11	50	23	10	90	25	9	129	25
2 nd Term	10	45	27	11	94	26	10	137	26
2011									
1 st Term	13	51	25	10	81	26	11	137	28
2 nd Term	11	48	24	11	92	25	10	128	26
2012									
1 st Term	13	43	26	10	70	24	11	118	30
2 nd Term	11	38	27	11	74	26	9	95	31
2013									
1 st Term	13	36	29	10	44	28	11	78	32
2 nd Term	11	27	33	11	53	31	8	61	27

This table reports the number of courses, the number of groups, and the average class size in each of the three academic programs by year and semester.

Table 4
Descriptive statistics of students, by degree program

	Finance & Accounting	Economics	Business Administration	Total
Female students (%)	52.78	47.12	53.3	51.27
Students from outside Madrid (%)	16.59	23.03	21.75	21.27
Students entering from standard nationwide examination (%)	96.83	97.2	96.84	96.95
Average entry score	7.481	8.047	8.689	8.278
	(1.591)	(1.859)	(1.694)	(1.791)

This table contains the percentage of students who are either females, from outside Madrid, or entering from the standard nation-wide exam procedure, and it also gives the average entry score. The average entry score is calculated over a maximum of 14 points, and the standard deviations are provided in parentheses. These statistics are given by academic programs, and for the total of students choosing the three programs employed in this research.

Table 5
Overall satisfaction, by department, and year

		Economics	Business	Engineering: BMD	Statistics	Econ. History
Year 1	2008	3.497	4.014	3.694	3.522	4.282
	2009	3.673	3.974	3.607	3.583	4.033
	2010	3.690	4.050	4.213	3.770	3.874
	2011	3.652	3.914	3.558	3.596	3.891
	2012	3.638	3.968	4.203	3.693	3.874
	2013	3.657	3.958	3.907	3.941	3.904
Year 2	2009					
	2010					
	2011	3.641	3.531		3.913	
	2012	3.615	3.597		3.791	
	2013	3.920	3.570		3.577	
Year 3	2010	3.919	3.769		3.507	
	2011	3.966	3.760		3.461	
	2012					
	2013					
Year 4	2011	3.518	3.506		3.027	
	2012	3.830	3.741		2.934	
	2013	3.815	3.646		2.768	
	AVERAGE	3.725	3.786	4.034	3.482	3.977
	STD. DEV	0.147	0.181	0.345	0.357	0.161

This table reports the overall satisfaction of students by years and departments teaching in any of the three degree programs employed in this research. These numbers are based on the questionnaires used by the university to evaluate overall student satisfaction with teachers on a scale between one (very negative) and five (very positive).

Table 6
Variable descriptions and descriptive statistics of the main regression variables employed

Variable Description	Mean	Standard Deviation	Min	Max
<i>Finance & Accounting</i> : a dummy variable that indicates whether the group belongs to this degree program.	0.172	0.377	1	
<i>Business Administration</i> : a dummy variable that indicates whether the group belongs to this degree program.	0.512	0.500	0	1
<i>Economics</i> : a dummy variable that indicates whether the group belongs to this degree program.	0.317	0.465	0	1
<i>Location of the program</i> : a variable that takes the value 1 if the campus is located in Getafe and 0 otherwise	0.906	0.292	0	1
<i>Teaching language: English</i> : a dummy variable that indicates whether the course is taught in English or not.	0.224	0.417	0	1
<i>PhD</i> : a dummy variable for each teacher in a given group.	0.479	0.499	0	1
<i>Average entry test score</i> : the average score in each group. The highest value this variable can have is 14.	8.288	1.396	5.801	11.62
<i>Proportion of females</i> : in each group.	0.512	0.112	0	0.875
<i>Proportion of non-academic Schools</i> : in each group.	0.031	0.037	0	0.272
<i>Class Size</i> : average number of students in each group.	30.16	8.786	6	84
<i>Proportion from outside Madrid</i> : average number of students in each group.	0.212	0.090	0	0.517
<i>Average age of the group</i> : for students	23.48	1.680	19.37	27.26
<i>Age</i> : for teachers.	42.94	8.846	23.89	70.43
<i>Research</i> : points obtained for research activity by the teacher of each group. The highest value this variable can have is 80.	14.82	14.50	0	80
<i>Top publications</i> : points obtained for top publications. The highest value this variable can have is 60. All teachers holding a PhD can access this ranking.	2.724	8.100	0	60
<i>Teaching activities</i> : points obtained for teaching activities. The highest value this variable can take is 50. All teachers holding a PhD can access this ranking.	9.601	16.16	0	47.72

This table contains a brief description of the variables employed in the regression used in the paper to analyze the relation between teaching quality and academic research. It also reports the precise way in which these variables are measured by University officials. The last four columns report the mean, standard deviation, and minimum and maximum values of the variables employed throughout the paper.

Table 7
Estimates of teaching quality, controlling for student and group characteristics

	Min	Max	Standard deviation	N
Overall	-10.979	7.832	1.667	1,955
Degree Programs				
Fin. & Acc.	-10.979	6.000	1.744	364
Economics	-7.476	5.749	1.612	603
Bus. Admin.	-6.044	7.832	1.673	988
Professor Category				
External	-7.476	7.832	1.683	1,195
Tenure track	-6.400	6.000	1.694	514
Associate	-10.979	4.653	1.601	203
Professor	-2.324	2.358	1.138	43
Year				
1 st year	-10.979	6.000	1.714	1,079
2 nd year	-5.379	7.832	1.619	769
3 rd year	-5.117	4.007	1.452	107
Department				
Business	-10.979	7.832	1.683	644
Economics	-6.258	5.749	1.602	900
Statistics	-6.166	6.622	1.758	244
Economic History	-6.400	4.259	1.785	128
Engineering: BMD	-6.626	3.203	1.772	39

This table contains descriptive statistics of the residuals estimated from OLS heterokedastic-robust standard error regression of the 1,955 alphas (conditional means of future grades of students) on the group characteristics. In particular, the table reports the descriptive statistics of the residuals from regression (2). Note that these residuals are the conditional means of future grades purged from students and group characteristics. These residuals are interpreted as our measure of teaching quality.

Table 8
Determinants of conditional means of future grades across all degree programs, controlling for group and professor characteristics with quadratic teaching experience

<i>Independent Variables</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>
FinanceAcc&Spanish&Getafe	2.561 (1.36)	4.050* (2.06)	4.441* (2.09)	4.564* (2.17)	4.433* (2.08)	4.537* (2.15)
FinancAcc. & English & Getafe	2.315 (1.24)	3.808* (1.97)	4.190* (2.00)	4.285* (2.07)	4.195* (1.99)	4.268* (2.06)
Economics & Spanish & Getafe	2.328 (1.26)	3.767+ (1.96)	4.175* (2.00)	4.254* (2.06)	4.175* (2.00)	4.231* (2.05)
Economics & English & Getafe	2.127 (1.14)	3.564+ (1.84)	3.959+ (1.90)	4.047+ (1.96)	3.965+ (1.89)	4.030+ (1.95)
Business & Spanish & Colmenar	3.169+ (1.73)	4.628* (2.43)	5.026* (2.43)	5.109* (2.50)	5.025* (2.42)	5.087* (2.48)
Business & Spanish & Getafe	2.778 (1.47)	4.250* (2.16)	4.635* (2.18)	4.711* (2.24)	4.628* (2.17)	4.685* (2.22)
Business & English & Getafe	2.652 (1.39)	4.115* (2.09)	4.517* (2.13)	4.589* (2.19)	4.508* (2.12)	4.561* (2.17)
Average entry test score	-0.156* (-2.28)	-0.218** (-2.85)	-0.215** (-2.81)	-0.205** (-2.65)	-0.212** (-2.77)	-0.204** (-2.63)
Proportion of females	-0.310 (-0.86)	-0.385 (-1.06)	-0.396 (-1.09)	-0.359 (-0.98)	-0.395 (-1.09)	-0.357 (-0.97)
Non-academic high	0.951 (0.89)	1.070 (0.99)	1.182 (1.11)	1.228 (1.15)	1.213 (1.14)	1.250 (1.17)
Class size	-0.000623 (-0.11)	0.00368 (0.66)	0.00311 (0.56)	0.00159 (0.29)	0.00302 (0.54)	0.00154 (0.28)
Outside Madrid	0.797+ (1.66)	0.929+ (1.95)	0.948* (1.98)	0.942* (1.99)	0.920+ (1.92)	0.917+ (1.93)
Average age of the group	-0.0696 (-1.21)	-0.118+ (-1.92)	-0.120+ (-1.94)	-0.135* (-2.18)	-0.120+ (-1.93)	-0.135* (-2.16)
Age (Teacher)			-0.0193 (-0.53)	-0.00855 (-0.23)	-0.0203 (-0.55)	-0.00940 (-0.25)
Square Age (Teacher)			0.000256 (0.66)	0.000129 (0.33)	0.000274 (0.70)	0.000144 (0.37)
Research			-0.000783 (-0.23)	-0.00129 (-0.35)	-0.00129 (-0.37)	-0.00175 (-0.47)
Top publications			0.0134+ (1.96)	0.0133+ (1.83)	0.0308* (2.06)	0.0284+ (1.79)
Square Top publications					-0.000492 (-1.51)	-0.00043 (-1.21)
Teaching Activities			-0.00416 (-1.36)	-0.00309 (-1.01)	-0.00539+ (-1.70)	-0.00414 (-1.30)
Professor dummies			YES	YES	YES	YES
Department dummies	NO	NO	NO	YES	NO	YES
Term dummies	NO	YES	YES	YES	YES	YES
Observations	1,955	1,955	1,955	1,955	1,955	1,955
R-squared	0.058	0.071	0.073	0.078	0.074	0.078
Log-likelihood function	-3786.2	-3772.9	-3770.7	-3765.8	-3770.1	-3765.4
F	8.612	7.064	5.810	5.527	5.895	5.423
Degrees of freedom	13	18	24	28	25	29
p-values	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Root MSE	1.684	1.675	1.675	1.673	1.675	1.673

This table contains the empirical results from OLS heterokedastic-robust standard error regressions of conditional means of future grades on group and professor characteristics, including research and teaching activities. In particular, these are the coefficients estimated from the regression given by equation (3). The observations of the dependent variable are weighted by the inverse of the standard error of the estimated alphas in the panel regression given by equation (1). The first two columns control for academic programs and locations, and for student and group characteristics. The last three columns also control for teachers' characteristics including different measures of research and teaching activities. We control for non-linear

teaching experience using the square of the teachers' age. Standard errors clustered at teacher level. Statistical significance by p -values reported in parentheses: + p -value < 0.10; * p -value < 0.05; ** p -value < 0.01; *** p -value < 0.01.

Table 9
Determinants of conditional means of future grades across all degree programs, controlling for group and professor characteristics with non-linear teaching experience

<i>Independent Variables</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>
FinanceAcc&Spanish&Getafe	2.561 (1.36)	4.050* (2.06)	4.165* (2.05)	4.459* (2.21)	4.121* (2.02)	4.403* (2.17)
FinancAcc. & English & Getafe	2.315 (1.24)	3.808* (1.97)	3.897+ (1.95)	4.167* (2.10)	3.862+ (1.93)	4.119* (2.07)
Economics &Spanish & Getafe	2.328 (1.26)	3.767+ (1.96)	3.883+ (1.96)	4.137* (2.10)	3.845+ (1.93)	4.084* (2.06)
Economics &English & Getafe	2.127 (1.14)	3.564+ (1.84)	3.663+ (1.84)	3.926* (1.98)	3.630+ (1.82)	3.879+ (1.96)
Business & Spanish &Colmenar	3.169+ (1.73)	4.628* (2.43)	4.748* (2.41)	5.002* (2.55)	4.708* (2.38)	4.949* (2.52)
Business & Spanish & Getafe	2.778 (1.47)	4.250* (2.16)	4.354* (2.15)	4.601* (2.28)	4.310* (2.12)	4.545* (2.25)
Business &English & Getafe	2.652 (1.39)	4.115* (2.09)	4.226* (2.09)	4.471* (2.22)	4.180* (2.06)	4.413* (2.19)
Average entry test score	-0.156* (-2.28)	-0.218** (-2.85)	-0.215** (-2.82)	-0.205** (-2.65)	-0.212** (-2.78)	-0.204** (-2.64)
Proportion of females	-0.310 (-0.86)	-0.385 (-1.06)	-0.403 (-1.11)	-0.362 (-0.99)	-0.403 (-1.11)	-0.361 (-0.98)
Non-academic high	0.951 (0.89)	1.070 (0.99)	1.158 (1.09)	1.214 (1.13)	1.182 (1.11)	1.232 (1.15)
Class size	-0.000623 (-0.11)	0.00368 (0.66)	0.00308 (0.55)	0.00155 (0.28)	0.00302 (0.54)	0.00150 (0.27)
Outside Madrid	0.797+ (1.66)	0.929+ (1.95)	0.954* (2.00)	0.946* (2.00)	0.931+ (1.94)	0.924+ (1.95)
Average age of the group	-0.0696 (-1.21)	-0.118+ (-1.92)	-0.121+ (-1.96)	-0.136* (-2.20)	-0.121+ (-1.95)	-0.136* (-2.18)
Non-Linear Age (Teacher)			-0.000524 (-0.05)	0.000372 (0.04)	0.000297 (0.03)	0.00111 (0.11)
Research			-0.0000800 (-0.02)	-0.000809 (-0.23)	-0.000478 (-0.14)	-0.00118 (-0.32)
Top publications			0.0126+ (1.83)	0.0127+ (1.76)	0.0278+ (1.89)	0.0263+ (1.69)
Square Top publications					-0.000435 (-1.34)	0.00039 (-1.11)
Teaching Activities			-0.00438 (-1.43)	-0.00321 (-1.04)	-0.00551+ (-1.72)	-0.00419 (-1.30)
Professor dummies			YES	YES	YES	YES
Department dummies	NO	NO	NO	YES	NO	YES
Term dummies	NO	YES	YES	YES	YES	YES
Observations	1,955	1,955	1,955	1,955	1,955	1,955
R-squared	0.058	0.071	0.072	0.077	0.073	0.078
Log-likelihood function	-3786.2	-3772.9	-3771.2	-3766.1	-3770.7	-3765.7
F	8.612	7.064	6.074	5.712	6.099	5.566
Degrees of freedom	13	18	23	27	24	28
p-values	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Root MSE	1.684	1.675	1.675	1.673	1.675	1.673

This table contains the empirical results from OLS heterokedastic-robust standard error regressions of conditional means of future grades on group and professor characteristics, including research and teaching activities. In particular, these are the coefficients estimated from the regression given by equation (3). The observations of the dependent variable are weighted by the inverse of the standard error of the estimated alphas in the panel regression given by equation (1). The first two columns control for academic programs and locations, and for student and group characteristics. The last three columns also control for teachers' characteristics including different measures of research and teaching activities. We control for non-linear teaching experience using the average age as the threshold, as shown in equation (4). Standard errors clustered at teacher level. Statistical significance by *p*-values reported in parentheses: + *p*-value < 0.10; * *p*-value < 0.05; ** *p*-value < 0.01; *** *p*-value < 0.01.

Table 10
Random allocation test for students

		Business Administration	Economics	Finance & Accounting	Total
Age	<i>F</i> Average	0.296	0.217	0.364	0.289
	<0.05	50	31	11	92
	<0.01	23	17	2	42
Non Ac. High School	χ^2 Average	0.667	0.741	0.697	0.705
	<0.05	0	0	0	0
	<0.01	0	0	0	0
Gender	χ^2 Average	0.407	0.305	0.393	0.374
	<0.05	21	17	9	47
	<0.01	9	3	3	15
Outside Madrid	χ^2 Average	0.447	0.408	0.607	0.479
	<0.05	22	17	3	42
	<0.01	9	1	0	10
Entry Test Score	<i>F</i> Average	0.322	0.147	0.372	0.284
	<0.05	61	46	21	128
	<0.01	30	35	8	73
Total # Tests		201	108	117	426

The reported statistics are derived from probit (Non Academic High schools, Gender and Outside Madrid characteristics) or OLS (Age and Entry test score characteristics) regressions of this observable student's characteristic on class dummies for each course in each degree program and academic year that we consider. The reported *p*-values refer to test of the null hypothesis that the coefficients on all the class dummies in each model are all jointly equal to zero. The test statistics are either a Chi-square or an F with varying parameters depending on the model. This column reports the average value of all tests. Columns labeled < 0.05 means number of courses for which the *p*-value of the test of joint significance of the class dummies is below 0.05. The same applies for < 0.01.

Table 11
Random allocation test for teachers: Regressions of teachers' observable characteristics on group characteristics

Panel A	<i>F</i> -statistic	<i>p</i> -value
Location of the program	7.40	0.00
Teaching Language: English	10.81	0.00
Average Entry test score	0.96	0.44
Average age of class	13.61	0.00
Proportion of females	1.32	0.25
Proportion from outside of Madrid	3.27	0.01
Class size	2.66	0.02
First year	8.71	0.00
Second year	6.17	0.00
Third year	3.39	0.01
Overall joint significance	10.64	0.00

The reported statistics are derived from a system of five seemingly unrelated simultaneous equations, where each observation is a group in a compulsory course for a total of 1,955 observations. The dependent variables are therefore the five teacher characteristics (research, top publications, teaching, age, and the category of teachers) and the independent variables are the eight group characteristics that generate perverse incentives among teachers when choosing a group. The reported *F*-statistics test the null hypothesis that the coefficients of each group characteristic are all jointly equal to zero in each equation of the system. The last row tests the null hypothesis that the coefficients on all the independent variables are jointly zero in all equations. All tests are distributed as $F(5;12,360)$, and the *F*-test of the last row is distributed as $F(45;12,355)$.

Panel B	Research Activities	Top Publications	Teaching Activities	Age	Associate Professor
Location of the program	2.11	3.72	3.09	-3.66	0.58
Teaching Language: English	-2.15	-0.35	-1.71	-7.07	-2.53
Average Entry test score	-1.23	0.39	-0.77	0.15	-0.07
Average age of class	-2.77	-3.12	-6.24	2.92	0.2
Proportion of females	-1.51	0.62	0.13	0.09	0.02
Proportion from outside of Madrid	-2.94	-3.12	-0.52	-1.6	-2.43
Class size	0.86	0.94	-1.77	-0.8	1.58
First year	-1.83	-5.71	-3.44	0.38	0.03
Second year	-1	-4.25	-1.82	-1.68	0.81
Third year	1.58	-2.05	-0.34	-0.01	-0.23
Intercept	3.04	2.94	5.75	4.41	0.24

The reported *t*-statistics are derived from the five individual regressions that are part of the seemingly unrelated system of simultaneous equations.

Table 12
Determinants of student satisfaction measured by questionnaires evaluating overall teaching quality

<i>Independent Variables</i>	<i>Estimated Coefficients (t-statistic)</i>	<i>Estimated Coefficients (t-statistic)</i>
Constant	1.121*** (13.44)	1.135*** (13.00)
Standardized Grades	0.00231 (0.64)	
Residual standardized grades		0.00202 (0.55)
Age	-0.00425*** (-11.18)	-0.00421*** (-10.87)
Satisfaction Evaluation Employed	0.823*** (172.39)	0.823*** (169.97)
Research	-0.00408*** (-11.84)	-0.00415*** (-11.84)
Top Publications	-0.00302*** (-7.11)	-0.00294*** (-6.81)
Teaching Activities	0.00373*** (18.23)	0.00377*** (18.06)
Finance Acc&Spanish& Getafe	-0.171* (-2.01)	-0.196* (-2.21)
Finance Acc. & English & Getafe	-0.308*** (-3.39)	-0.327*** (-3.47)
Economics &Spanish & Getafe	-0.124 (-1.35)	-0.148 (-1.56)
Economics &English & Getafe	-0.222* (-2.44)	-0.249** (-2.62)
Business & Spanish & Getafe	-0.0981 (-1.10)	-0.129 (-1.36)
Business & English & Getafe	-0.180* (-2.00)	-0.205* (-2.14)
Term Dummies	YES	YES
Department Dummies	YES	YES
Professor Category Dummies	YES	YES
Year Evaluation Dummies	YES	YES
Observations	58,721	56,933
R-squared	0.425	0.424
Log-likelihood function	-53,968.0	-52,373.8
F	1,595.4	1,556.7
df	26	26
Root MSE	0.607	0.607
Groups	1,955	1,955

This table reports the estimated coefficients from a fixed effects panel data model where the dependent variable is the overall satisfaction of the students obtained from anonymous questionnaires. The regressions control for academic programs and locations, and for teachers' characteristics including different measures of research an