Research Policy 46 (2017) 19–29

Contents lists available at ScienceDirect

Research Policy

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

Research paper

Research performance and teaching quality in the Spanish higher education system: Evidence from a medium-sized university

Joaquín Artés [a,*], Francisco Pedraja-Chaparro [b], Mª del Mar Salinas-JiméneZ [b]

[a] Departamento de Economía Aplicada IV, Universidad Complutense de Madrid, Ciudad Universitaria, S/N, 28040, Madrid, Spain
[b] Universidad de Extremadura, Spain

ARTICLE INFO

Article history:
Received 17 February 2016
Received in revised form
13 September 2016
Accepted 17 October 2016

JEL classification:
I20

Keywords:
Research output
Teaching performance
Higher education

ABSTRACT

This paper studies the relationship between research performance and teaching quality in the context of the Spanish university system. We investigate whether there is a relationship between being an active researcher and teaching quality of college professors in Spain. We use a data set from the University of Extremadura, which contains information on teaching evaluations and research performance over a ten year period (from 2001–2002 to 2011–2012). Our results suggest that, on average, professors who are more involved in research obtain better results in their teaching evaluations. We also suggest that this positive link between research and teaching is non-linear, as we find a larger improvement in teaching quality from additional research at lower levels of research intensity. Additionally, we show that the relationship between teaching and research is not constant along the distribution of teaching scores, and that the teaching quality of professors in the lower quantiles is much more related to their research intensity than that of professors in the top quantiles.

1. Introduction

This study analyses the teaching-research nexus in the context of the Spanish university system. In most countries there is a shared view that universities should perform both research and teaching activities, which in the aggregate are seen as complements. Reasonable arguments for both positive and negative effects of research on teaching can be found in the literature. The complementary view is usually based on the idea that research may create positive spillovers on teaching by facilitating up-to-date courses and a deeper understanding of the relevant topics. On the contrary, these activities could also be thought of as being substitutes if one considers constraints of time, effort and funding allocation (Marsh and Hattie, 2002). We can therefore hypothesize different relationships between research and teaching rather than a single link, with these relationships depending on contextual factors such as the type and level of research, the academic discipline, or the level and the mode of delivering of teaching (Brew, 1999).

The debate on the relationship between teaching and research has a long tradition among academics and brings to the forefront some relevant issues for university authorities and public policy in regards to how these activities relate, the optimal mix between teaching and research and the incentives put in place to improve the quality of both teaching and research conducted by universities. Whether research and teaching are complements or substitutes may motivate increased funding for one or the other activity. If research contributes to improving the quality of teaching, this positive external effect of research would provide an argument for increasing the funding devoted to research activities. In a similar way, the net effect of research on teaching may also motivate the need for professors to integrate these activities or to specialize in one of them, thus having important implications at the organizational level of universities and departments. As noted by Hughes (2004), a positive relationship between research and teaching would motivate locating these activities closely together whereas a non-significant relationship (or a negative one) would support the idea of separating them into research-only and teaching-only institutions.[1] Finally, a crucial point to promote quality in both teaching and research refers to the incentive schemes, which could affect the allocation of time and effort to these activities. Although incentives to teaching are often established, most incen-

* Corresponding author.
E-mail address: jartes@ucm.es (J. Artés).

[1] This discussion goes back to the nineties (see, for example, Elton, 1992), but there is still an open debate, at least in Europe, on the convenience (or not) of driving research and teaching further apart (see, for example, Dosi et al., 2006; or Karagiannis, 2009).

http://dx.doi.org/10.1016/j.respol.2016.10.003
0048-7333/

tives schemes are based on research output and this may bias the optimal balance between teaching and research (Sylos Labini and Zinovyeva, 2008). If the emphasis to achieve career progression is mainly placed on the quality of research, many academics could regard teaching duties as a “necessary evil” (Karagiannis, 2009), thus neglecting their teaching activities or allocating less time and effort to them.[2]

Universities in Spain, as is common around the world, have the double mission of teaching and contributing to knowledge through research.[3] These two activities are generally seen as complementary and incentives schemes are set to enhance teaching and research quality, although greater emphasis for the academic career is placed on research. Most Spanish universities measure teaching quality using teaching evaluations based on students' perceptions. Although it is generally acknowledged that these subjective assessments do not directly measure learning outcomes and could be biased by students' expectations,[4] most universities rely on them to evaluate teaching, as there is no widely accepted alternative that is as easy to use on a regular basis (Alwood et al., 2015; Marsh, 2007). In the case of research performance, incentives are built mainly upon officially recognized research evaluations that are conducted by education authorities following an external review process. Academics in Spain can submit their research for evaluation every six years and a national committee evaluates the five most relevant contributions produced in that six-year period (*sexenio*) and decides to accord (or not) an official recognition of that research period. A positive evaluation of that research period implies salary increases, providing a research incentive.[5]

As in other countries, a potential concern in the Spanish system is that the focus on research may lead to a decrease in the teaching quality of Spanish universities. In this study, we use the measures of teaching quality (evaluations) and research performance (*sexenios*) typically used by the Spanish authorities to evaluate professors, and analyze whether there is a relationship between research and teaching during a ten year period (2001/2002 to 2011/2012) in the University of Extremadura, which is a medium-sized university located in the southwest of Spain.

Our detailed data and extended period of study allow us to contribute to the existing literature in several ways. First, although there is a large literature on the relationship between teaching and research (see, for example, Hattie and Marsh, 1996; Colbeck, 1998; or Marsh and Hattie, 2002) for Anglo-Saxon countries, the evidence is scarce for other countries, despite the fact that these countries have very different university systems.

Second, while most of the previous literature models the relationship between teaching and research as linear and find at most a small link, we allow for a flexible non-linear relationship and find a quantitatively relevant effect. This result is important because combined with the fact that most of the papers finding a small and mostly linear effect refer to Anglo-Saxon universities confirms that conclusions from those studies are not directly applicable to other university systems. According to our results the positive association between research and teaching is driven by the low teaching performance of professors of low research intensity in comparison with those of medium or high intensity, but we do not find a significant difference between the teaching scores of professors of medium and high research intensity.

Third, we provide evidence that the relationship between research and teaching is not constant across different quantiles of the distribution of teaching scores, being much stronger for lower quantiles. This is important because, by focusing on the mean effect, previous studies are unable to provide information on how research relates to teaching for professors of different levels of teaching quality. We have to be cautious in the interpretation of our results because our identification relies on all common variables that affect both being active in research and teaching effectiveness being controlled for in the regression. Nonetheless, unlike in previous studies, our focus across the distribution of teaching quality allows us to inform policy recommendations for the whole distribution of teachers. We find that, although the magnitude of the effect decreases along quantiles, research is positively associated with teaching at all levels of the distribution up to the 90th quantile. This implies that only for excellent teachers (those above the 90th threshold of the teaching quality distribution), research is not a significant explanatory factor of their teaching quality.

Fourth, our results are particularly relevant in the context of Spanish universities. Our study uses the same measures of research intensity and teaching quality that are currently being used by Spanish education authorities to evaluate research and teaching of all professors across Spain, which provide our results with immediate policy implications. We find that despite their limitations, these measures point to a significant non-linear association between research and teaching quality. Professors in our sample that have a medium or high level of research intensity over the period obtain, on average, better teaching evaluations than their less involved in research peers. To our knowledge there is only one published paper that has documented a similar relationship- García-Gallego et al. (2015), for the University Jaume I-.[6] We obtain our conclusions using an extended period of time, and data and measures of research and teaching that are more generalizable across universities. Therefore, an additional contribution of the study is to show that the positive link between research and teaching is not specific to a single university and data, and also that it appears when using more policy-ready measures of teaching and research.

Finally, we believe that the study of the link between teaching and research can contribute to the on-going debate regarding the reform of the university system in Spain. Since the passing of new regulations in 2001,[7] the system has moved towards a performance-based system,[8] a pillar of which was the creation of the National Agency for Quality Assessment and Accreditation (ANECA) in 2002. The new law changed the hiring and promotion process for faculty in public universities and a positive evaluation from this Agency (or similar regional agencies) is now required for individuals to be able to apply for a teaching position. The requirements for obtaining a positive evaluation are, however, subject to debate as experts disagree on the role that research should play in recruiting, salaries and promotions.[9] While research plays a cen-

[2] Mas-Colell (2003) offers a good discussion on the incentives schemes to teaching and research and compares two extremes situations to achieve a given teaching-research mix: “The institution can choose first a high teaching talent... and then rely on incentives to reach the desired research level. Or it can focus first on research talent and rely on the incentive part to guarantee the teaching objective”. Although “the idea to choose the academic staff mainly by its research potential is controversial in Europe and it is less practiced than in the USA”, the second alternative seems superior to him.

[3] See Sánchez-Barrioluengo (2014) for a discussion of Spanish Universities not only as centers of excellence in teaching and research but also on third mission.

[4] See McPherson (2006) on the determinants of students' assessments of teaching. Furthermore, the issue of whether student's evaluations really reflect effective learning or teaching quality has recently been questioned in several works (see, for example, Beleche et al., 2012; Galbraith et al., 2012; or Braga et al., 2014).

[5] See Jiménez-Contreras et al. (2003) or Osuna et al. (2011) for the role of this research indicator in the Spanish university context.

[6] There is also a recent working paper by Rodriguez and Rubio (2013) for the University Carlos III, in Madrid that reaches similar conclusions.

[7] The law that regulates the university system is called *Ley Organica de Universidades*, and was approved on December 6th, 2001.

[8] For a discussion on performance-based university systems, see the work by Hicks (2012).

[9] For example, a recent report written by a committee of independent experts appointed by the Spanish government to suggest policies to improve

tral role in promotions, the concern has been raised that recruiting and promoting professors based on mostly on their research ability could potentially be detrimental for the teaching quality of Spanish universities.[10] As there is scarce empirical evidence on the topic, an important additional contribution of this paper is to inform such policy debate.

The structure of the remainder of the paper is as follows. Section 2 reviews the literature on the teaching-research nexus and presents the existing empirical evidence. Section 3 offers an overview of the institutional framework in Spain. Section 4 presents the empirical model and section 5 provides a description of the data used. Section 6 presents the results. Finally, the paper closes with a discussion of the results in Section 7 and a summary of the main conclusions in Section 8.

2. The teaching-research nexus: theory and empirical evidence

Newman (1992) coined the term "teaching-research nexus" to refer to the links between research and teaching. Different manifestations of the relationship between teaching and research may appear, within and between disciplines, depending on whether teaching/learning is conceived either as the transmission of knowledge or as the process of organizing and generating that knowledge (see, for example, Brew, 2003). In either case, several arguments have been raised in the literature to support the positive effects of research on learning. When teaching is seen as the transmission of knowledge, it is generally highlighted that research helps teachers in mastering current developments in their discipline and they may, consequently, teach more up-to-date courses and promote a deeper understanding of the relevant topics. However, one could also think that research tends to be too specialized to enter into undergraduate courses and this could lead researchers either to offer courses at a too high of a level or to distort the curriculum toward their own research in detriment of a broader study program (Karagiannis, 2009). On the other hand, it has since long been argued that the process of scientific inquiry is the central organizing concept of learning, so researchers are better placed to motivate students, to develop attitudes of inquiry and to enhance research skills in students (Hattie and Marsh, 1996).

At the instructor level, a positive impact of research on teaching has also been motivated by the common abilities underlying both research and teaching, specifically the values and skills that lead to excellence in research (e.g. dedication, organization, originality, or critical thinking) are also likely to lead to excellence in teaching (e.g. knowledge of the subject, planning and presentation of the courses, or clarity of course objectives and requirements). Nevertheless, some authors highlight that different abilities and skills are required to perform both activities, suggesting that the personality characteristics of supportiveness, tolerance and warmth tend to be positively correlated with effective teaching whereas they appear to be negatively related to research productivity (Feldman, 1987). Moreover, it has also been argued that research has a negative impact on teaching, mainly because of a trade-off in time and effort spent on each of these activities. In addition, when faculty careers depend on research, it provides incentives to reduce the time and effort spent on teaching, so that allocation of time and effort would be biased in favor of research activities (Marsh and Hattie, 2002; Karagiannis, 2009).

Given that the literature on the research-teaching nexus offers arguments both for positive and negative effects of research on teaching, the question of whether these activities are complements or substitutes, in the aggregate, becomes an empirical issue. Early work in the late eighties and early nineties suggest that the overall correlation between teaching and research is close to zero, although slightly positive (see the meta-analyses by Feldman, 1987; Allen, 1996; or Hattie and Marsh, 1996). Nevertheless, the diversity of results across studies makes the empirical evidence inconclusive. Although many empirical works have analyzed the link between teaching and research, empirical studies greatly differ in the way they measure research and teaching activities and in the scope of the analysis (see Verburgh and Lindblom-Ylanne, 2007, for a critical review of the existing empirical evidence). This heterogeneity in the variables used in analysis and in the ways to measure them leads to very different results across studies. Moreover, comparable datasets across universities in different countries (or even within a given country) are not publicly available, so the results obtained are difficult to generalize and are often specific to a single university or department, or even to a specific discipline in a single institution.

Recent work on the teaching-research nexus, while it continues to suffer from a reliance on narrow dataset, has advanced in separating the effects at the individual and departmental (or institutional) levels, in considering the possibility that non-linearities may exist, and in widening the scope of variables under analysis. Complementarities between research and teaching may exist at the departmental (or university) level even when these activities appear to be not related, or negatively related, at the professor level. This would lead to an internal specialization where the department (or university) provides high quality teaching and research but some academics are specialized in research whereas others are involved in teaching activities (Coate et al., 2001; Gautier and Wauthy, 2007). It could also be the case that the combination of complementary relationships and the constraints of time and effort give rise to a non-linear relationship between teaching and research, so assuming a linear relationship, as was done in previous studies, would reduce the magnitude and significance of observed correlations. Several empirical works tend to support this view, pointing to a positive effect of research on teaching up to a threshold level, but once this level is reached, increasing research efforts would reduce teaching performance (García-Gallego et al., 2015; Mitchell and Rebne, 1995; Stack, 2003). Furthermore, the empirical work on the teaching-research nexus has broaden the outcomes of interest by considering students' related variables, such as students' performance in the labor market (Urwin and Di Pietro, 2005; Sylos Labini and Zinovyeva, 2008), or teachers' related variables, such as type of contract or tenure (Bettinger and Long, 2010; Figlio et al., 2013). In all cases, the recent empirical evidence suggests that teaching and research activities are not independent, but that a positive (often non-linear) relationship appears between them.

Most of the existing empirical work studies Anglo-Saxon countries, whereas studies in continental Europe are relatively scarce. One notable exception is the work by Sylos Labini and Zinovyeva (2008), who work with a rich data set from Italian Universities and find that students' satisfaction with teaching positively correlates with department-level indicators of academic research quality, as measured by expert evaluation scores and bibliographic indicators. In the case of Spain, little empirical work has been carried out on the links between research and teaching. Nevertheless, two studies have recently analyzed this relationship, suggesting in both cases, that a positive relationship holds between them. García-Gallego et al. (2015) use data on 604 university professors at the University Jaume I, in Castellón, over the period 2002–2006. This data cover various disciplines: humanities, social sciences, economics,

the university system recommends to base recruiting and promotions mostly on research. This recommendation gave rise to a heated policy debate. The report can be downloaded from the webpage of the Spanish Ministry of Education: http://www.mecd.gob.es/prensa-mecd/dms/mecd/servicios-al-ciudadano-mecd/participacion-publica/sistemauniversitario/propuestas-reforma.pdf.

[10] See for example page 33 of the expert report mentioned in the previous footnote.

management, natural sciences, and engineering. They find a significant non-linear and positive effect of research on teaching quality. Finally, the working paper by Rodriguez and Rubio (2013) analyze the relationship between teaching quality and research productivity at University Carlos III, in Madrid. Their data range from the academic years 2008/2009 to 2011/2012 and cover the fields of Business Administration, Economics, and Finance and Accounting. Using value-added measures of students' performance they find a positive and significant relationship between high levels of research and teaching quality.

In sum, although the empirical evidence tends to be specific to single universities or to specific disciplines, the results of recent literature on the relationship between teaching and research suggest that, in the aggregate, these activities are complements, so the positive effects of research on teaching tend to offset the possible negative impact derived from time and effort constraints, at least up to a certain threshold.

In this paper, we widen the number of non-Anglo-Saxon cases for which a positive relation between research and teaching is documented and explore further the non-linearities of such a relation. In addition our paper is novel in two aspects. First, one limitation of previous analyses is the focus on mean effects. Most previous studies use OLS regressions to analyze the link between teaching and research. OLS regressions allow for inferences at the mean of the dependent variable, in this case, at the mean of the teaching quality distribution. While inferences at the mean are interesting, it is often the extremes of the distribution that are of policy concern. In our case, for example it is interesting to know whether there is diversity in the way research is associated to teaching quality along the distribution of teachers and whether the teaching quality of the top professors in the sample, for example, is related to research in a similar way than that of professors in lower quantiles, as this could lead to different incentive schemes. We estimate a quantile regression model that shows that the relation of research to teaching quality differs significantly across quantiles and that is non-existent for the top professors in the sample.

Second, another limitation of previous studies, particularly for the case of Spain is that their results do not allow for direct policy recommendations as they use indicators of research that are quite detailed but are also university specific, which makes results difficult to use by authorities at the national level. As explained in the introduction, in Spain, as in other European countries, all professors are subject to the same regulations and incentive system, which are set at the national government. The main salary incentive to do research comes from the *sexenios*, a research evaluation that determines every six years if a professor has performed enough research during the period (both quantitatively and qualitatively) to accrue the evaluation and the salary incentive that comes with it. The *sexenios* have become a fundamental part of the Spanish university system and are the main tool used by the Spanish government to assess the research activity of Spanish professors. While the *sexenios* are widely recognized as a successful tool to evaluate research, the concern has been raised that the focus on hiring and promoting based on the research activities measured by the *sexenios* may lead to a reduction of teaching quality. Our paper contributes to this debate by focusing on the relationship between *sexenios* and teaching evaluations, thus providing results with immediate policy implications.

3. Institutional framework

In Spain, as of 2013 there were 50 public universities and 33 private universities. We focus on public universities because all of them combine teaching and research activities, which is not the case for all private universities. The 50 public universities are geographically distributed across Spain according to population and organizational needs. There is slightly less than one public university per every one million people and at least one public university in each of the 17 Spanish regions, although some of the most populated areas have many more public universities (e.g. Madrid has 5 and Catalonia has 6). In this study we analyze data from the University of Extremadura which is the only public university in the region of Extremadura. As of 2013, the University of Extremadura had around 24,000 students and 2000 professors, making it a medium-size University in the Spanish context.[11] In terms of research, the University of Extremadura is also placed in a medium range, with a relative position of 29 in the Scimago Country Rank in 2013. It offers classes and degrees in a wide range of fields and attracts students mostly from cities and towns within the region of Extremadura.

All public universities in Spain, including the University of Extremadura, are subject to the same regulations regarding the hiring and promotion of professors. There are several different types of contracts under which a professor could be hired to teach at a public university. Each contract has different requirements and implies different wages. The main distinction is whether the contract is tenured or non-tenured. In this paper we focus on civil servant tenured contracts because only professors holding these contracts (e.g. *Profesor Titular* or *Catedrático*) can apply for the research evaluations that we use to construct our independent variable of interest. Typically, professors can be promoted to a tenure contract after holding an Assistant Professor non-tenured contract (*Profesor Ayudante Doctor*) during a few years. To be promoted to a tenured contract professors' have to obtain an external positive tenure evaluation performed at the national level by an experts committee (ANECA). A few years after a professor has been promoted to a standard Associate Professor Contract (*Profesor Titular*),[12] another research evaluation will determine if she can be promoted to Full Professor (*Catedrático*).[13]

The salaries of civil servant tenured professors in all Spanish public universities are determined by two parts. The fixed part depends on the rank of the professor and is higher for Full Professors than for Associates and higher for Associates than for non-tenured faculty. The variable part of the wage has two parts, one of them depends on experience, and results in a wage raise every three (*trienios*) and five years (*quinquenios*).

In addition, every six years professors holding a civil servant tenured contract can apply for another wage increase based on their research output (*sexenio*). Contrary to the *quinquenios* and *trienios*, the *sexenios* are not awarded to every professor automatically. A national committee, *the National Committee for the Evaluation of Research Activity* (CNEAI),[14] evaluates each application according

[11] A variety of statistics about the university system in Spain can be obtained from the web page of the Spanish Ministry of Education: http://www.mecd.gob.es/educacion-mecd/areas-educacion/universidades/estadisticas-informes.html Additionally, a recent ranking of Spanish universities developed by the Fundacion BBVA and the IVIE can be obtained from: http://www.u-ranking.es/analisis.php.

[12] There are other contractual figures in the Spanish system that offer conditions similar to a tenure contract. For example a contract of Contratado Doctor is also a permanent contract and requires professors to obtain an ANECA evaluation before applying for this contract. However the research requirements to obtain this contract are lower and so are some of the benefits associated with the contract. In addition this is not a civil servant contract which prevents professors holding this contract from participating in standard research evaluation calls. This is one of the reasons why we focus only on professors holding standard Associate and Full Professors Contracts.

[13] The research requirements that determine promotions have substantially changed over the last decades, which means that professors with very different research productivities may hold the same contracts.

[14] CNAI is different from ANECA. CNAI evaluates research activity for the purposes of accruing wage increases for tenured professors. ANECA performs an overall evaluation of professors for the purposes of tenure promotions that includes research and teaching activities as well as other activities such as administrative service.

to a set of guidelines that takes into account the quality of the publications, and only those professors who are evaluated positively receive the wage increase. Applying for the *sexenio* is voluntary and aproximately 12% of professors that could apply, have never applied for a *sexenio*. The research evaluations are considered relatively rigorous and many professors are denied the *sexenio.* According to recent CNEAI statistics, 60% of Associate Professors and 30% of Catedráticos are denied the *sexenio* on average.[15]

In order to obtain a *sexenio*, professors need to submit their top five research contributions published during the six year period that they want to be evaluated. The six years to be evaluated do not need to be consecutive. For example a less active professor who has a first contribution in 1991, another in 1997 and three others in 2001, 2002 and 2008, could choose to be evaluated for any six specific years between 1991 and 2008, and obtain one *sexenio* for the whole period. However, a more active researcher with many contributions in each of the years could apply during the same period for up to three *sexenios* (one every six years). This is a useful feature of the Spanish system because it allows us to classify professors working in the same university according to their different research intensity. This provides useful variation that we exploit in our identification strategy.

4. Empirical model

Our main empirical strategy consists on regressing a measure of teaching quality on research performance, controlling for observable differences between researchers and students. In particular we run a series of Ordinary Least Squares (OLS) regressions of the following type:

$$\text{Teaching Quality}_{it} = b_0 + b_1\text{Research Intensity}_i + b_2 \text{Professor Characteristics}_{it} + b_3\text{Student characteristics}_{it} + \varepsilon_{it} \quad (1)$$

Where *Teaching quality* is a measure of teaching quality of each professor, such as student evaluations; *Research Intensity* is a measure of how active the professor is in research. *Professor Characteristics* are observable differences between professors such as age, qualifications, the teaching load or the field of expertise; and *Student characteristics* are aggregates of the socio-demographic characteristics of the students taught by each professor. The coefficient of interest would be $b_{1.}$ Assuming that the common variables that affect being active in research and teaching effectiveness are observable and controlled for in the regression, b_1 captures the effect of being an active researcher on teaching effectiveness. In all the different specifications of this model the standard errors are clustered at the professor level to account for potential within-group correlation.

An alternative estimation strategy could have been to account for professor's idiosyncratic characteristics and for time effects in a fixed-effects panel data model. However, such a model would be problematic given our research variable. As explained in more detail in the next section, we measure research intensity as the number of relative research evaluations (number of positive research evaluations over the number of potential research evaluations) that each professor has at the end of the study period, in 2012. The advantage of using the number of relative research evaluations is that it is the most widely used measure to evaluate research intensity for policy purposes in Spain. The main disadvantage is that we cannot use this variable in a fixed effects model. This is because a fixed effects specification uses within-variation, which means that our identification would arise in such a model from changes in the research intensity of professors over time. A change in a professor's recorded number of relative research evaluations from one year to another does not capture changes in research activity during that year, but changes over the six year period for which the research evaluation is accrued. For this reason in our case is preferable to measure research intensity for all professors in the sample on the last year of data and use between-variation to identify the effect. Therefore, given our set of controls, our identification arises from comparing professors that have different research performance over the period of study controlling for professors' and students' characteristics.

We also estimate a quantile regression model. This model is useful for our purposes because it allows us to test whether the effects of research intensity are different for professors of different teaching ability. The OLS specifications do not allow for testing this because they estimate the effect of the variable of interest at the conditional mean of the dependent variable. Instead, the quantile regression model measures the impact of an endogenous variable at different points of the conditional distribution of the dependent variable. This approach allows us to estimate the effect of research intensity at different points of the teaching quality distribution. In particular, we estimate the following standard quantile regression model (Koenker and Basset, 1978):

$$Teaching\ Quality_{it} = Xi\beta_q + \mathrm{u}_{\theta i}, \text{ with} \mathrm{Quant}_\theta(\text{Teaching Quality}_{it}|\mathrm{X_i}) = Xi\beta_\theta \quad (2)$$

Where X_i includes the same variables as the right hand side of Eq. (2), and β_θ is the vector of parameters to be estimated. Quant_θ $(\mathrm{TeachingQuality_i} \mid \mathrm{X_i})$ refers to the θth conditional quantile of *TeachingQuality* given X. The vector of parameters is found by solving the following minimization problem:

$$\min_{\beta \in R^K} \sum_i (Teaching\ Quality_{it} - X_{it}\beta_\theta) \quad (3)$$

where ρθ (ε) is a check function defined as ρθ (ε) = θ ε if ε > 0 or ρθ (ε) = (θ-1) if ε > 0. The set of coefficients β that solve the minimization problem [4] is found using standard linear programming techniques. The set of coefficients will be different for each quantile to be estimated. Note that the quantile regression model reduces the risk of misspecification due to making incorrect assumptions about the conditional distribution over the predicted values for the quantiles. As these assumptions cannot be used either to determine the size of standard errors, we compute those by bootstrapping. In our case we used one thousand replications in the bootstrap routine.

5. Data

We estimate Eqs. (1) and (2) using data provided by the University of Extremadura. Our database contains the teaching evaluations, the research performance and several other characteristics of all the professors of the university during the ten year period between 2001/2002 and 2011/2012.[16]

5.1. Measures of research intensity

For each professor in our database we measure their research productivity as the ratio between the number of *Sexenios* they actually have in 2012 and the maximum number of *Sexenios* they could

[15] This number refers to the 2009 call. See the CNEAI website for the full report: http://www.mecd.gob.es/ministeriomecd/organizacion/organismos/cneai/memorias-informes.html. For a comprehensive evaluation of success rates on the sexenio applications see Jiménez-Contreras et al. (2003).

[16] The University of Extremadura provided a version of the database in which names of professors were substituted with an id number to preserve anonymity.

have had. We calculate the maximum possible number of *sexenios* from the number of *trienios* (three year periods of work under contract at the university), so that for every two *trienios* we infer that a professor could have had one *sexenio*. This measure takes into account that even though the first year that a professor is able to apply for a *sexenio* is the first year holding a civil servant tenured contract, the first research evaluation can be obtained for earlier years as long as the professor has held a contract. To see how this variable works in practice, we can use the same example that we used in the previous section, in which we had a less active researcher publishing six papers in eighteen years and a more active researcher publishing many papers in each of the years. The less active researcher, would have a research intensity of 1/3 because this professor obtained one *Sexenio* out of a maximum of three. The more active researcher would have a research intensity of 1 because she obtained three *Sexenios* out of a maximum of three. An advantage of this variable for our purposes is that many universities in Spain are currently using either the same or very similar measures to determine research productivity.

In our benchmark specification we enter our measure of research intensity linearly. However, as it is possible that the relation between teaching quality and research intensity is non-linear, we also estimate the basic model using a non-linear version of the research intensity variable. One possibility to estimate non-linearities is to add a quadratic polynomial on research intensity. Such a specification however is problematic in our case for two reasons. On one hand, our measure of research intensity is censored at 0 and at 1, which creates boundary problems. Additionally, a quadratic specification imposes a specific parametric form to the relation between teaching and research. For this reason, we prefer to estimate non-linearities in a flexible way using different vectors of categorical variables constructed from the research intensity variable. Specifically, in our preferred specification we create a categorical variable with three values, 0, 1 and 2 according to whether the value of the continuous research intensity variable is either less than 1/3, between 1/3 and 2/3, or more than 2/3. The idea is to classify research intensity in three groups, low, medium and high research intensity and test whether teaching evaluations' results differ across research intensity groups. As the cutoff points to define the research intensity groups are arbitrary, we also estimate a model using other cutoffs and groupings. Specifically, in the results section we show results of a specification in which we divide researchers into four groups of research intensity instead of three (with cutoffs at 0.25, 0.5 and 0.75), which allow us to test the robustness of results and to gain additional insight on how research intensity is associated with teaching evaluations.

5.2. Measuring teaching quality

We measure teaching quality using the summary of the teaching evaluations of each professor in each class and each year. The summary is constructed by the University of Extremadura using student's responses to a teaching evaluation questionnaire. The teaching questionnaire is quite standard and includes a mix of questions related to clarity of lectures, punctuality, complementary material provided by each instructor, and the ability to motivate students, as well as an overall teaching effectiveness question such as *"what is your overall satisfaction with the teaching ability of this instructor?"*.[17] The university assigns a weight to each question and constructs an overall index of teaching performance for each professor in each class. This index was provided to us by the University of Extremadura in a scale from 0 to 10, where 0 stands for "very bad" and 10 stands for "very good". The teaching evaluations were filled in by students anonymously, in-class, during the last weeks of the semester and before the final grades were released.

5.3. Control variables

We include as control variables the experience, the rank and the field of study of the professor, the degree in which the class is taught, the number of classes a professor is teaching each term and whether the class is elective or a higher division class. *Experience* is measured as the number of *trienios* (number of three year periods that a professor has accumulated in the public university system) and is included as a vector of dummy variables to account for non-linearities.[18] The experience dummies capture differences due to experience and also to professors' unobserved characteristics such as motivation or enthusiasm that are associated with age. The *rank* of the professor (Full or Associate) is defined as a dummy that takes value 1 for Full Professors and 0 for Associates and captures potential differences in teaching across professors of both categories that may exist due to the different requirements needed to hold each position. The variable *field of study* includes 147 different fields. Apart from controlling for differences among professors, the field dummies also capture differences among students. Differences among the characteristics of students pursuing different degrees in different fields can also be captured by a vector of dummies (*degree*) reflecting the degree in which the class is taught, which is a variable with 192 categories. For example, *Principles of Economics* is taught to political science students, to engineering students, to students of economics and to students pursuing several other degrees. In the University of Extremadura, as in many other Spanish universities, students pursuing the same degree are grouped together in the same class (unlike for example in U.S. universities in which the same class may have students pursuing different degrees). Therefore, including the degree in which the class is taught controls for student characteristics. We could potentially include both the field dummies (147 categories) and the degree dummies (192 categories) in the same model. However we run into multicollinearity problems when both sets of dummies are entered jointly because both variables capture a mixture of student and professor characteristics. To avoid this problem, we grouped professors into 5 broader fields (Health Sciences, Experimental Sciences, Social Sciences, Arts and Humanities and Engineering and Architecture) and estimated the model including this grouped variable jointly with the vector of 192 ◦ dummies. Together these two vectors of categorical variables account for differences among students taking classes in similar degrees, and for characteristics that are common among professors of similar fields (e.g. difficulty of classes or to some extent different ability levels). To account for the possibility of professors of higher research intensity selecting themselves into more specialized or more difficult classes, we also include two dummy variables: *Upper Level* and *Elective.* The first variable, *Upper Level,* measures whether the class is taught in the later years within the curriculum and captures that more specialized classes are usually taught in those later years. The second variable, *Elective,* measures whether the class is elective or not. This variable controls for the fact that elective classes are more specialized, and are offered by professors and chosen by students based on their own interests. Finally, we also include a linear time trend and year dummies. The *linear trend* is defined as the year in which the teaching evaluation is answered and is included to control for

[17] A copy of the questionnaire can be obtained from authors upon request.

[18] Our database also includes the specific age of each professor. As this variable is highly correlated with *experience*, we do not include both *age* and *experience* in the regression to avoid potential multicollinearity problems.

Table 1
Summary Statistics.

	Tenured Professors			
	Mean	Standard Deviation	Min	Max
Teaching Evaluations	7.03	1.57	0	10
Research Evaluations	1.90	1.42	0	6
Research Intensity	0.57	0.37	0	1
Research 1 Low	0.28	0.45	0	1
Research 1 Medium	0.31	0.46	0	1
Research 1 High	0.41	0.49	0	1
Research 2 Low	0.28	0.45	0	1
Research 2 Medium Low	0.22	0.41	0	1
Research 2 Medium High	0.16	0.36	0	1
Research 2 High	0.35	0.48	0	1
Elective	0.23	0.42	0	1
Upper division	0.30	0.46	0	1
Trienios	7.65	2.90	2	14
Full Professors (Catedrático)	0.20	0.40	0	1
Associate Professors (Titular)	0.80	0.40	0	1
Health	0.20	0.40	0	1
Science	0.30	0.46	0	1
Social Sciences	0.23	0.42	0	1
Arts and Humanities	0.18	0.39	0	1
Engeneering and Arquitecture	0.09	0.28	0	1
Number of Observations	**5015**			

the increase in average teaching evaluation scores over the years.[19] Time dummies are included to capture shocks that are common across observations in each year.[20]

The unit of observation is the teaching evaluation, which means that for each professor we can have several observations corresponding to the same year. The estimation sample of civil servant tenured professors for which we have all the variables consists of 5015 observations corresponding to 708 tenured professors. On average we observe 7.08 evaluations per tenured professor in the sample.

6. Results

We start our presentation of the results with a description of the summary statistics of the main variables.

In Table 1 we show the descriptive statistics. Columns 1 to 4 show the mean, the standard deviation, the minimum and the maximum values for the estimating sample of civil-servant tenured professors. These columns show that approximately 20 per cent of our sample consists of Full Professors while 80 per cent are Associates. The average experience of professors in the samaple is 7.65 *trienios* (the average age is 51.4 years). The average number of positive research evaluations (*sexenios*) is 1.90. Average research intensity (number of positive evaluations divided by the maximum possible number of positive research evaluations) is 0.57, while the average teaching score is 7.03 in a scale of 0 to 10. This table also shows that our sample includes professors from all fields of knowledge. The most represented areas of knowledge in our sample are Sciences and Social Sciences, with more than 20% of the observations belonging to each of these categories, while the field of the lowest representation is Engineering and Architecture, with 9% of observations.

Table 2 shows the distribution of teaching scores by research intensity. In panel A we show results of the categorical variable that classifies researchers in three groups depending on whether their research productivity is lower than 1/3, between 1/3 and 2/3, or higher than 2/3. This panel shows that people of medium or high research intensity obtain better teaching evaluations than those of low research intensity. A *t*-test of the differences in means shows that these differences are statistically significant. The finding that professors with a medium or high level of research intensity obtain better teaching scores than professors with low research activity holds for all fields, although the differences between groups, for some pairs, are not statistically significant. This panel also shows that when comparing the group of medium research intensity with that of high research intensity, differences are quantitatively small, and in many cases also non-significant. This result points to a potential non-linearity in the relationship between teaching and research: the difference in teaching evaluations seems to be quite pronounced when comparing professors of low research intensity with professors of medium and high intensity, and appears to be less pronounced or inexistent when comparing professors of medium and high research intensity. Panel B in Table 2 shows that the conclusion holds when we classify professors in four groups (with cutoffs at ¼ and ¾) according to their research intensity. In this case it is again clear that professors in the lower group of research intensity obtain on average lower scores on their teaching evaluations than professors in the other three groups. On the other hand, when comparing professors in the medium-low, medium-high and high research intensity, differences among them are again quantitatively smaller and in many cases statistically non-significant.

The mean comparisons of Table 2 are suggestive of a relationship between research intensity and teaching performance but they have to be interpreted with caution because there may be large differences in observable or unobservable characteristics between researchers and professors with high and low research activity. In Table 3 we show the results of different versions of the models of Eq. (1), in which some of these differences are accounted for. Columns 1 is estimated using the continuous definition of research intensity. The results of column 1 show that the coefficient of this variable is statistically significant and has a value of 0.3495, which would be the marginal effect on teaching scores of moving from not doing any research to being an active researcher. To interpret this coefficient correctly, it is worth putting it in the context of a specific example. According to the results, researchers with, for example, a research productivity of 2/3, have teaching scores 0.12 points higher than researchers with a research productivity of 1/3, or approximately 1.7%. While this result may seem quantitatively small, it is worth noting that the standard deviation for the teaching score variable is only 1.57 (see descriptive statistics of Table 1), which means that small quantitative changes in teaching scores do imply significant increases in teaching quality. In particular, a 0.12 increase in teaching scores represents approximately 7.6% of the standard deviation.

In column 2 we use the non-linear definition of research intensity in which we classify professors into three groups according to their research intensity. The coefficient of the medium research intensity group has a value of 0.3163. This implies that, compared to the group of low research intensity, professors of medium research intensity obtain teaching evaluations that are 0.3163 points higher on a scale of 0 to 10, or 20.14% of the standard deviation. Additionally, if we compare the coefficients of the medium and high research intensity groups, we see that they are much more similar (the coefficient of the high intensity group is 0.3795, or 24.17% of the standard deviation), and we cannot disregard that they are equal. This result points to research intensity having a significant and economically relevant non-linear effect on teaching quality. The implication is that while having low research activity clearly leads to lower teaching evaluations, there seems to be only a small difference between being a medium or a high intensity researcher.

[19] There is a positive trend in teaching scores in the university during the period of study. For example the average teaching scores went from 6.42 during the academic year 2002–2003 to 7.28 in the last year for which we have data, which is 2011–2012.

[20] We checked that the inclusion of all the covariates in the model did not cause multicollinearity problems using the variance-inflation factor (VIF). Results of this test as well as the correlation matrix are available under request.

Table 2
Teaching Scores by research intensity.

	All	Health Sciences	Experimental Sciences	Social Sciences	Arts and Humanities	Engeneering and Arquitecture
Panel A: Research 1						
Rsearch 1 Low	6.65	6.19	6.74	6.62	7.11	6.26
Research 1 Medium	7.06	7.41	7.04	7.07	7.33	6.49
Research 1 High	7.23	7.46	6.95	7.05	7.51	7.08
t-tests	*a,a,b*	*a,a,ns*	*a,a,ns*	*a,a,ns*	*c,a,c*	*ns,a,a*
Panel B: Research 2						
Rsearch 2 Low	6.66	6.19	6.73	6.65	7.08	6.08
Research 2 Medium Low	7.05	7.37	6.96	6.82	7.61	6.30
Research 2 Medium High	7.08	7.25	7.01	7.45	7.03	6.41
Research 2 High	7.24	7.40	7.04	6.80	7.47	7.06
t-tests	*a,a,a,ns,b,b*	*a,a,a,ns,c,a*	*a,a,a,ns,ns,ns*	*a,a,a,b,ns,ns*	*a,ns,a,ns,ns,a*	*ns,ns,a,ns,a,a*

Note: Each cell represents the mean of the teaching evaluations of all the professors in the estimating sample (Titulares and Catedráticos) that belong to each category. The teaching evaluations are measured in scale from 0 to 10. The *t*-tests row represents the significance of a *t*-test of the differences in means between each group. In panel A, each entry on the *t*-test row represents the *t*-test of: (Low vs Medium, Low vs High, Medium vs High). In panel b each entry of the *t*-test row refers to: (Low vs Medium Low, Low vs Medium High, Low vs High, Medium Low vs Medium High, Medium Low vs High, Medium High vs High). Significance levels of *t*-tests: a = 1%, b = 5%, c = 10%, ns = not significant.

Table 3
Regression Results. Dependent Variable: Teaching Scores.

	1	2	3
Variables	OLS	OLS	OLS
Research intensity	0.3495**		
	[0.174]		
Research 1 (Medium)		0.3163**	
		[0.144]	
Research 1 (High)		0.3795**	
		[0.158]	
Research 2 (Medium-Low)			0.3132**
			[0.146]
Research 2 (Medium-High)			0.3020*
			[0.177]
Research 2 (High)			0.4188**
			[0.172]
Elective	0.8490***	0.8448***	0.8445***
	[0.076]	[0.077]	[0.076]
Upper division	−0.1183	−0.1212	−0.1184
	[0.096]	[0.095]	[0.096]
Teaching Load	−0.0171	−0.0158	−0.0166
	[0.036]	[0.036]	[0.036]
Rank (Titular)	−0.017	0.002	0.0136
	[0.142]	[0.140]	[0.142]
Linear trend	0.0823***	0.0806***	0.0810***
	[0.016]	[0.016]	[0.016]
Experimental Sciences	−0.4346***	−0.4206***	−0.4172***
	[0.155]	[0.156]	[0.156]
Social Sciences	−0.6859**	−0.6622*	−0.6562*
	[0.347]	[0.365]	[0.361]
Arts and Humanities	−0.3987	−0.4182	−0.4044
	[0.248]	[0.254]	[0.252]
Engeneering and Arquitecture	−1.0035***	−0.9856***	−0.9804***
	[0.301]	[0.303]	[0.304]
Constant	6.7625***	6.7358***	6.6634***
	[0.864]	[0.866]	[0.886]
Year dummies	Yes	Yes	Yes
Degree dummies	Yes	Yes	Yes
Experience dummies (trienios)	Yes	Yes	Yes
Observations	5015	5015	5015
R-squared	0.201	0.203	0.203

Clustered (by professor) standard errors in brackets.
*** p < 0.01.
** p < 0.05.
* p < 0.1.

Column 3 confirms that this result is robust to using the four-group categorization of the research intensity variable (the coefficient of the medium-high category is, in this case, non-significant, but has a similar magnitude). These findings are consistent with theories that argue that the complementarities between research and teaching exist up to a certain threshold, but once that threshold is reached, additional research intensity creates smaller additional benefits for teaching. In addition, and contrary to other papers such as Mitchell and Rebne (1995) or Stack (2003), we do not find any detrimental effects on teaching of being a high intensity researcher compared to a medium intensity one.

Note that our results cannot be explained by professors of high research intensity selecting themselves into elective or more specialized classes because those factors are controlled for by the variables *Elective* and *Upper Division*. Our results, however, and despite the relatively large set of covariates included in the empirical model can be explained by different causal mechanisms. One of them is that research and teaching are complements so doing research allows professors to improve their teaching. On the other hand, it is possible that the correlation found between being an active researcher and teaching quality is due to selection in terms of ability or motivation. If this is the case, the positive correlation would not be explained by the complementarities between teaching and research, but, instead, it would be the consequence of individuals with higher teaching ability having also characteristics that allow them to do research better (e.g. overall ability or motivation). Therefore, in terms of policy, the implication that can be derived from this model is that researchers that have at least a medium level of research will, on average, obtain better teaching scores than less research-active peers. Whether this is due to differences in ability or complementarities cannot be inferred from our results.

A second related limitation of the results of Table 3 is that it is possible that the relationship between research and teaching is not the same for teachers of different innate teaching ability. This is relevant because in terms of policy, many times we are interested in understanding the effects of a policy incentive for people located at different points of the distribution, for example for better teachers or for worse teachers. Table 4 shows the results of the quantile regression model. To save space we only show the results when our preferred measure of research intensity –the three group categorical variable- is used as the independent variable of interest.[21] Note that in this table we are estimating the effects of research intensity along the distribution of teaching scores. In other words, we want to check if research intensity exhibits a similar association with teaching quality for professors at different points of the teaching ability distribution.

[21] Estimation of the quantile model using either the continous measure of research intensity or the four-group variable lead to the same conclusions and are available upon request from the authors.

Table 4
Quantile Regression Results. Dependent Variable: Teaching Scores.

	1	2	3	4	5
Variables	q10	q25	q50	q75	q90
Research 1 (Medium)	0.4487***	0.4869***	0.4276***	0.1964***	0.0789
	[0.148]	[0.121]	[0.083]	[0.075]	[0.076]
Research 1 (High)	0.6630***	0.6551***	0.3774***	0.2109***	0.1152
	[0.174]	[0.124]	[0.091]	[0.079]	[0.081]
Elective	1.0449***	0.9864***	0.7919***	0.5316***	0.4110***
	[0.130]	[0.093]	[0.065]	[0.058]	[0.060]
Upper division	0.3246**	0.1400*	0.1281**	0.0558	0.065
	[0.128]	[0.085]	[0.064]	[0.056]	[0.065]
Teaching Load	−0.0041	−0.0116	−0.0145	−0.0024	−0.0089
	[0.049]	[0.031]	[0.022]	[0.019]	[0.019]
Rank (Titular)	0.157	0.0134	−0.1137	−0.1382*	−0.1409*
	[0.159]	[0.116]	[0.085]	[0.084]	[0.077]
Linear trend	0.1052***	0.0814***	0.0689***	0.0576***	0.0631***
	[0.025]	[0.018]	[0.011]	[0.011]	[0.010]
Experimental Sciences	−0.2471	−0.3551***	−0.2308**	−0.1243	−0.0596
	[0.156]	[0.108]	[0.092]	[0.082]	[0.082]
Social Sciences	−0.6372***	−0.4026***	−0.2662***	−0.1137	0.0056
	[0.196]	[0.137]	[0.103]	[0.087]	[0.092]
Arts and Humanities	−0.1217	0.0918	0.2934***	0.4630***	0.4202***
	[0.207]	[0.132]	[0.102]	[0.081]	[0.082]
Engeneering and Arquitecture	−1.3178***	−1.1035***	−0.6929***	−0.3587***	−0.3077**
	[0.310]	[0.219]	[0.162]	[0.115]	[0.127]
Constant	3.8008**	6.8483***	6.7989***	7.1910***	8.0056***
	[1.508]	[1.405]	[1.020]	[0.832]	[0.726]
Year dummies	Yes	Yes	Yes	Yes	Yes
Degree dummies	Yes	Yes	Yes	Yes	Yes
Experience dummies (trienios)	Yes	Yes	Yes	Yes	Yes
Observations	5015				

Bootstrapped (1000 replications) standard errors in brackets
*** $p<0.01$.
** $p<0.05$.
* $p<0.1$.

The table shows several interesting results. First, the coefficients of the research intensity variables are always positive and statistically significant up to the 90th quantile. This suggests, again, that doing research is associated with better teaching regardless of the individuals' innate ability for teaching: professors at the bottom (e.g. 25th quntile) and the top of the teaching distribution (e.g. 75th quantile) that are active researchers obtain better teaching evaluations than peers in the same part of the distribution with lower levels of research. Second, the coefficient of the research intensity variable varies significantly along the teaching scores distribution. Research has a much stronger association with better teaching in the lower part of the conditional teaching scores distribution than in the upper part of the distribution. In particular, the coefficient of high research intensity for professors of lower teaching ability (10th and 25th quantile) is as high as 0.4487 for the 10th quantile and 0.4869 for the 25th quantile. This implies that among individuals of lower teaching ability, those with at least medium research intensity obtain around 0.45 points higher scores on their teaching evaluations compared to individuals in the lower group of research intensity, or 28.6% of the standard deviation. For professors of high research intensity the coefficient goes up to 0.66 for the 10% quantile and 0.65 for the 25% quantile, or more than 40% of the standard deviation. On the other hand, professors on the upper part of the teaching scores distribution obtain much more similar teaching scores regardless of their research intensities. In this case the coefficient of the medium research intensity is 0.1964 (12.5% of the standard deviation) for those in the 75th quantile and 0.08 (5% of the standard deviation) for those on the 90th quantile. A *t*-test of the differences between the coefficients of the lower and upper quantiles (e.g. 25th quantile vs 75th quantile) confirms that these differences are statistically significant at the 1% level.

A third interesting result is that the relationship between teaching and research changes along the teaching distribution. For the lower quantiles there is a strong and almost linear positive relationship between teaching and research. For the median quantile there is still a positive relationship between being a researcher of at least medium intensity and teaching, but additional research does not make a difference anymore. Finally in the upper quantiles, the non-linearities are still present but the difference between being at least a medium intensity researcher compared to a low-intensity researcher is quite small and even non-significant for the 90th quantile.

In sum, the implications of the quantile model are: 1) among the teachers in the lower part of the teaching distribution, those with medium research intensity do significantly better than professors with low research intensity, and those with high research intensity do better than those with medium research intensity; 2) among professors located around the mean of the teaching distribution professors of low research intensity do significantly worse than their peers in terms of teaching, but there is no significant difference between having a medium or a high level of research intensity; 3) among the better performing teachers in the sample, research does not show a strong association with teaching scores. We now discuss the policy implication of these results.

7. Discussion and policy implications

In terms of policy, a first implication of our results relates to the hiring of new faculty. In Spain and in many universities across the world, the access to a tenured position requires demonstrating a certain level of research. The research requirements to access such a position in Spain have been increasing over the last few years in accordance to international patterns. Provided that our regression framework adequately controls for common variables that affect being active in research and teaching effectiveness, our results suggest that increasing research requirements would not result in an overall worsening of teaching quality over time if the new cohorts of professors are required to show at least medium levels of research. This would be because requiring an active research agenda could attract individuals of higher overall ability for teaching, and because, as shown by the results of the quantile model, for those individuals of similar teaching ability, research is also associated with better teaching.

A second implication comes also from our quantile results. Only at the very top of the distribution of teaching quality (above the 90th percentile), professors conducting a moderate level of research do not obtain better teaching scores. Acknowledging that caution is granted when making recommendations based on a model that requires relatively strong identifying assumptions, this finding suggests that promotions based on research will likely lead to overall improvements on teaching quality of Spanish universities. According to our results promotions based solely on teaching quality should only be granted to the very top teachers as measured by objective teaching quality measures. Our results suggest a system that combines hiring and promotions based on requiring at least a medium level of research as a general rule, with teaching incentives in terms of wage increases and awards for excellent teachers (i.e. those above the 90th percentile of the teaching distribution). Based on our evidence, such a system is likely to provide both adequate incentives to do research and the same time improve overall teaching quality.

Additionally, our results, combined with recent evidence for the Spanish case (García-Gallego et al., 2015; Rodriguez and Rubio, 2013), show that relationship between research and teaching in Spain is quantitatively non-neglegible and non-linear. This finding departs from the conclusions of earlier research, summarized in reviews such as Feldman (1987), Hattie and Marsh (1996) or Marsh and Hattie (2002), which suggests that the effect of research on teaching quality is small or non-existent for Anglo-Saxon countries. Therefore, our paper shows that inferences based on Anglo-Saxon systems are not directly applicable to countries with substantially different higher education systems such as Spain. Our results however, can be reconciled with previous evidence by taking into account that the sample of tenured professors working in an average Spanish university like the one used in our study exhibits a higher degree of heterogeneity in research intensity compared to average Anglo-Saxon universities typically analyzed in previous studies. If we perform the analysis making our sample more homogeneous by only keeping in the sample professors of at least medium-low research intensity (e.g. those with research intensities greater than ¼), we would find only a small and non-significant association between research intensity and teaching quality, and our results would look much more similar to most of the previous studies analyzed in Marsh and Hattie (2002).

Finally, it is important to emphasize that previous studies base their conclusions on models that make inferences at the mean of the distribution of teaching quality. Given that the policy debate is many times focused on how to create incentives at different points of the distribution of teaching (e.g. how to reward very good teachers or how to prevent bad teachers to enter the system), it is important to know how teaching is associated with research at all points of the distribution. Our results show that the positive association between both activities appears at all points of the distribution except for teachers in the top ten of the distribution of teaching quality, but even among those, higher research intensity is not negatively associated with teaching. We also find that the magnitude of the association changes substantially along quantiles, and that the differences in teaching quality are particularly large for the lower quantiles. Therefore, we do not find any evidence for an often cited critique that those professors doing research make worse teachers.

8. Conclusion

This paper finds that there is a positive correlation between research performance and the teaching scores obtained by college professors teaching in a Spanish medium sized university. In particular, this paper finds that professors of a medium or high level of research intensity obtain significantly better results on teaching evaluations compared to professors of low research intensity. We also find that the result is non-linear, as professors with high research intensity do not obtain better teaching evaluations than professors of medium research intensity. In addition, our quantile regression model shows that the positive and significant correlation exists along all the quantiles of the conditional teaching scores distribution. The quantile model, however, also shows that the magnitude is much larger for professors in the lower parts of the teaching scores distribution. This implies that being a moderately active researcher is more highly associated with being a high-quality teacher at the lower part of the teaching distribution than in the top quantiles.

Our results could be explained by two different causal mechanisms. On the one hand the positive relationship between research and teaching is consistent with theories that argue that doing research allows professors to be up to date with their fields and that this is translated into better teaching. On the other hand, our results could be explained by innate individual ability. If some of the characteristics that make a good researcher are also helpful for being a good teacher (e.g. organization, creativity, motivation, technical knowledge, etc), a positive association between teaching and research activities points to individuals that have those characteristics being able to do both activities better.

Although our sample does not allow for distinguishing the mechanisms and, therefore, should be interpreted with caution, our results still have direct policy implications regarding the hiring of new faculty. At the general level, a first decision to be made about the organization of a university system is to opt for a system in which many universities are teaching-only or for a system in which universities combine both activities. Our results suggest that combining both activities in the same institution is likely to be beneficial for teaching quality. In addition our results have policy implications in terms of the organization of an academic career. Recruiting systems based on at least a moderate level of research productivity are not likely to result in a decrease in teaching quality. Our results show that only in the case of professors of very high innate teaching ability (those on the top 10% of the distribution), doing at least a moderate level of research would not lead to significant improvements of their teaching quality. Thus, a system that selects professors of demonstrated ability to perform research would be unlikely to have a negative effect on teaching quality compared to a system in which selection is based solely on teaching. On the other hand, the policy implications of our results regarding promotions depend on the mechanism that explains the association. If our results stem mostly from complementarities between research and teaching effectiveness, promotions based on research would lead to better teaching effectiveness. However, the model does not allow us to make this claim if the mechanism that is behind the results is differences in unobserved ability.

Finally, our paper has limitations that provide opportunities for future research. On one hand, future research should further investigate the causal mechanisms that explain the relationship between research and teaching. Disentangling whether the positive correlation is due to selection or to complementarities is important for improving selection in hiring and promotion and for the design of career incentives. On the other hand, as many universities use similar methods for the evaluation of teaching quality and research output, the construction of comprehensive databases that include many universities and years of study should be a priority of this research agenda, so that results can be further generalized. Our paper used measures of teaching and research quality that are relevant for policy purposes because they are similar to the most commonly used indicators across Spanish universities, and in addition the have the added advantage of their simplicity. However, future research should further investigate the best way to evaluate teaching and research outputs in a manner that keeps the simplicity of the indicators currently used but improves their accuracy to capture teaching and research quality. In this regard, the construction of homogeneous measures of research productivity that incorporate more time variation would allow for the use of econometric models better able to disentangle the causal mechanisms underlying the positive correlations between teaching and research. Universities and policy authorities should make the construction of such comprehensive databases a priority so that policy decisions can be made after a careful consideration of all available evidence.

References

Allen, M., 1996. Research productivity and positive teaching evaluations: examining the relationship using a meta-analysis. J. Assoc. Commun. Admin. 2, 77–96.

Alwood, S., Wallstad, W.B., Siegfried, J.J., 2015. Teaching economics to undergraduates. J. Econ. Lit. 53, 285–325.

Beleche, T., Fairris, D., Marks, M., 2012. Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. Econ. Educ. Rev. 31, 709–719.

Bettinger, E.P., Long, B.T., 2010. Does cheaper mean better? The impact of using adjunct instructors on student outcomes. Rev. Econ. Stat. 92, 598–613.

Braga, M., Paccagnella, M., Pellizzari, M., 2014. Evaluating students' evaluations of professors. Econ. Educ. Rev. 41, 71–88.

Brew, A., 1999. Research and teaching: changing relationships in a changing context. Stud. Higher Educ. 24, 291–301.

Brew, A., 2003. Teaching and research: new relationships and their implications for inquiry-based teaching and learning in higher education. Higher Educ. Res. Dev. 22, 3–18.

Coate, K., Barnett, R., Williams, G., 2001. Relationships between teaching and research in higher education in England. Higher Educ. Q. 55, 158–174.

Colbeck, C.L., 1998. Merging in a seamless blend: how faculty integrate teaching and research? J. Higher Educ. 69, 647–671.

Dosi, G., Llerena, P., Sylos Labini, M., 2006. The relationships between science, technology and their industrial explotation: an illustration through the myths and realities of the so-called 'European paradox'. Res. Policy 35, 1450–1464.

Elton, L., 1992. Research, teaching and scholarship in an expanding higher education system. Higher Educ. Q. 46, 252–268.

Feldman, K., 1987. Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: a review and exploration. Res. Higher Educ. 26, 227–298.

Figlio, D., Shapiro, M., Soter, K., 2013. Are tenure track professors better teachers? In: Working Paper WP 13–18. Institute for Policy Research, Northwestern University.

Galbraith, C.S., Merrill, G.B., Kline, D.M., 2012. Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in Business related classes? A neural network and Bayesian analyses. Res. Higher Educ. 53, 353–374.

García-Gallego, A., Georgantzís, N., Martín-Montaner, J., Pérez-Amaral, T., 2015. (How) Do research and administrative duties affect university professorsí teaching? Appl. Econ. 47 (45), 4868–4883.

Gautier, A., Wauthy, X., 2007. Teaching versus research: a multi-tasking approach to multi-department universities. Eur. Econ. Rev. 51, 273–295.

Hattie, J., Marsh, H.W., 1996. The relationship between research and teaching: a meta-analysis. Rev. Educ. Res. 66, 507–542.

Hicks, D., 2012. Performance-based university research funding systems. Res. Policy 41, 251–261.

Hughes, M., 2004. The relationships between research and teaching in higher education – a review of the literature 1990–2002. In: Ocasional Paper. Brighton Business School, University of Brighton.

Jiménez-Contreras, E., de Moya Anegón, F., Delgado López-Cózar, E., 2003. The evolution of research activity in Spain: the impact of the National Commission for the Evaluation of Research Activity (CNEAI). Res. Policy 32, 123–1142.

Karagiannis, S.N., 2009. The conflicts between science research and teaching in higher education: an academic's perspective. Int. J. Teach. Learn. Higher Educ. 21, 75–83.

Marsh, H.W., Hattie, J., 2002. The relation between research productivity and teaching effectiveness: complementary, antagonistic or independent constructs? J. Higher Educ. 75, 603–641.

Marsh, H.W., 2007. Student's evaluations of University teaching: dimensionality, reliablitiy, validity, potential biases and usefulness. In: Perry, J.C. (Ed.), The Scholarship of Teaching and Learning in Higher Education: an Evidence-based Perspective. , pp. 319–383.

Mas-Colell, A., 2003. The European Space of higher education: incentive and governance issues. Rivista di Politica Economica, 9–27, November–December.

McPherson, M., 2006. Determinants of how students evaluate teachers. J. Econ. Educ. 37, 3–20.

Mitchell, J.E., Rebne, D.S., 1995. Non-linear effects of teaching and consulting on academic research productivity. Socioecon. Plann. Sci. 29, 47–57.

Newman, R., 1992. Perceptions of the teaching-research nexus: a framework for analysis. Higher Educ. 19, 351–375.

Osuna, C., Cruz-Castro, L., Sanz-Menendez, L., 2011. Overturning some assumptions about the effects of evaluation systems on publication performance. Scientometrics 86, 575–592.

Rodriguez, R., Rubio, G., 2013. Teaching Quality and Academic Research Paper Presented at the X Jornadas de Economía Laboral. www.uam.es/otros/jaeet13/comunicaciones/02_Rendimiento_de_la_educacion/Rodriguez-(linkbreak)_Rubio.pdf.

Sánchez-Barrioluengo, M., 2014. Articulating the three-missions of Spanish universities. Res. Policy 43, 1760–1773.

Stack, S., 2003. Research productivity and student evaluation of teaching in Social Science classes: a research note. Res. Higher Educ. 44, 539–556.

Sylos Labini, M., Zinovyeva, N., 2008. The relationship between academic research, teaching quality and graduates' employment outcomes. Mimeo www.hha.dk/nat/workshop/2008/nz0412. pdf.

Urwin, P., Di Pietro, G., 2005. The impact of research and teaching quality inputs on the employment outcomes of postgraduates. Higher Educ. Q. 59, 275–295.

Verburgh, J.E., Lindblom-Ylanne, S., 2007. Investigating the myth of the relationship between teaching and research in higher education: a review of empirical research. Stud. Philos. Educ. 26, 449–465.