Teacher quality and student skill acquisition. An analysis based on PIRLS-2011 outcomes

Jorge Calero and J. Oriol Escardíbul

University of Barcelona & IEB. Barcelona. Spain.

Corresponding author: Jorge Calero. Facultat d'Economia i Empresa, Avda. Diagonal 690, 08034 Barcelona. Email address: jorge.calero@ub.edu. Telephone number: 0034 934021812.

Biographical notes

Jorge Calero is tenured Professor of Applied Economics at the University of Barcelona, department of Economics. Researcher of the Barcelona Institute of Economics (Institut d'Economia de Barcelona). He is currently coordinator of the Interdisciplinary Research Group on Educational Policies GIPE-IGEP. Previous Head of the Council of the Educational System Evaluation (Catalonia). His main research interests are the economics of education, educational policies evaluation, equity issues in education, and economics of the welfare state.

Josep-Oriol Escardíbul is associate professor of Applied Economics at the Universitat de Barcelona. He is currently Vice-rector for Finance of the University of Barcelona. President of the Spanish Association of Economics of Education. Research fellow at the Barcelona Economics Institute (IEB) & GIPE.

Teacher quality and student skill acquisition. An analysis based on PIRLS-2011 outcomes

Abstract

This article examines the question of teacher quality and its effects on the skills acquisition of primary school students in Spain. We use an education production function in which we incorporate teachers' fixed effects, estimated by means of a multiple regression model. Specifically, we examine the acquisition of reading skills by drawing on data from PIRLS-2011. The results obtained allow us to conclude that teachers constitute an important input in the acquisition of reading skills. Additionally, we analyse which observed teachers' characteristics can be related to teacher quality.

By identifying the quality of each teacher in the sample, we are also able to determine where the "best" and "worst" teachers work. The results indicate that there is a greater probability of finding high quality teachers in privately owned schools and in schools where the students come from families with higher levels of economic and socio-cultural resources.

Keywords: teacher quality, skills acquisition, reading skills, PIRLS

JEL codes: H52, I25, I28

1. Introduction

The role played by teachers in the acquisition of cognitive skills has long been the object of attention, and of much controversy, in the economics of education and in education policy analysis. Indeed, the intuition that the teachers' role is central has not always been easy to demonstrate empirically. The factors that might potentially explain why certain teachers are "better" or "worse" in helping their students acquire cognitive skills have come under close scrutiny, but have not as yet been clearly established. At the same time, education policies, in recent years, have led to greater attention being

paid to teachers, not only in quantitative terms – with discussions about the effect of student-teacher ratios being especially intense – but also in qualitative terms.

It is against this backdrop, that we undertake this study, which seeks to answer the following three research questions: i) Do teachers matter in the skills acquisition of primary school students? ii) What factors determine the quality of teachers? iii) In which schools do the "best" and "worst" teachers work?

In order to answer these questions, we adopt a methodological approach based on the estimation of a production function using a multiple regression model in which we introduce teacher fixed effects. The explanatory variable in the production function is the score recorded by each student on the reading skills test. The two levels analyzed are, on the one hand, that of the student and, on the other, that of the school. Information about the characteristics of each specific teacher is linked to that of their students. This methodological approach – applied here for the first time to the Spanish case – is possible thanks to the availability of data from the Progress in International Reading Literacy Study (PIRLS-2011), conducted by the IEA, one of the few examples of an assessment tool that integrates information about student skills and teacher characteristics. The PIRLS assessment, focused specifically on reading literacy achievement, is administered to fourth grade students in primary education. The assessment has no academic repercussions, which favours the absence of contamination in the results and ensures an effective evaluation of the students' actual reading skills (Jacob and Levitt 2003).

In this study, we identify the fixed effect of each teacher in the production function as a proxy for their level of quality. Here, we use the term "quality" as shorthand to refer solely to the teachers' ability to increase the students' cognitive skills,

or, even more specifically, their reading skills. It says nothing about other areas in which we might speak of teacher quality and, likewise, it says nothing about their ability to develop the students' non-cognitive skills.

The methodology we employ, therefore, allows us to assign a level of quality (with the limitations that we describe above) to each teacher, depending on the level of skills acquired by their students. The relationship between quality levels and student outcomes will enable us to answer the first research question. Subsequently, by establishing a relationship between teacher quality and their observed characteristics, as recorded by PIRLS, we can address the second research question. The third research question (In which schools do the "best" and "worst" teachers work?) can also be answered with the information obtained about teacher quality.

The rest of the article is structured as follows. Section 2 undertakes a review of the relevant literature. Sections 3 and 4 are dedicated to the empirical approach employed: section 3 describes the data and section 4 the methodology used. Section 5 presents a descriptive analysis of the variables of interest for conducting our analysis. The results of the estimations that allow us to answer the first and second research questions are presented and discussed in section 6. Section 7 seeks to answer the third research question. Finally, section 8 presents the conclusions and discusses their implications for education policy.

2. Empirical evidence of the relevance of teachers in students' academic performance

Studies of the impact of school factors on student academic achievement gained a certain relevance following the release of the Coleman Report in the United States (see

Coleman et al 1966). In general, much of the empirical evidence tends to indicate that the effect of educational resources on student achievement is low, at least in the more developed countries (for a review, see Calero and Escardíbul 2015). In this section, we consider studies that focus their analysis on identifying the effects of teachers (and teacher characteristics) on student skill acquisition and academic achievement. In so doing, we do not consider quantitative aspects, that is, those related to the impact of the number of students per teacher or the number of students per class (class-size); rather, we focus on aspects related to the specific characteristics of the teachers.

The extensive bibliography on this specific field, mostly referring to the case of English-speaking countries, has allowed three main findings to be identified:

1) The quality of teachers is relevant in the process of skills acquisition.

2) The effect of the quality of the teacher on this process is heterogeneous: its magnitude depends on factors such as educational level and type of student.

3) The specific characteristics of the teachers account for a small proportion of the variation in the students' skills.

The empirical evidence that supports the above statements is outlined below.

1) Most of the research finds that the effects of the quality of teachers on skills acquisition vary between slight and moderate. Among the studies that provide evidence on these lines are those of Aaronson et al. (2007), Hanushek et al. (2005), Hanushek (2011), Kane et al. (2008), Leigh (2010), McCaffrey et al. (2009), Rivkin et al. (2005), Rockoff (2004), and the review undertaken by Canales and Maldonado (2018). These studies are based on education production functions, where the dependent variable (to be explained) is an approximation of the students' educational performance or their

skills acquisition, and among the explanatory variables are dichotomous variables (which are given a value of "1" for a given teacher and "0" for the rest). These variables represent the so-called fixed effects that capture the influence of one particular teacher (and not another) for each student. In fact, the impact of fixed effects is associated with the concept of teacher "quality". Studies of this type include in their analyses the students' prior attainment (usually as recorded in the previous academic year) as an explanatory variable, making it possible to analyse the value added between school years. The results obtained using this methodology reveal whether the fact that a student has one teacher or another matters, regardless of any concomitant characteristics. However, with this approach it is not possible to establish the causes of this relationship.

Specifically, a consistent finding in studies using this methodology, when applied to primary school, is that increasing teacher "quality" by one standard deviation translates into an increase of approximately 10-20% of a standard deviation of academic achievement. For the case of the UK, Slater et al (2012) report that this impact rises to 27.2%. There are other studies that compare the effect of the quality of the teachers with other input-based policies. For instance, Rivkin et al. (2005) show how the effect of a one-standard deviation improvement in teacher quality on student test scores is estimated to be larger than the effect of a ten student reduction in class size.

2) The heterogeneity of the effect can be appreciated when segmenting the samples by a diversity of factors (see, for example, the reviews made by Hanushek and Rivkin, 2012 and by Koedel et al., 2015). Among the most important factors is educational level. Lockwood and McCaffrey (2009) show that teacher effect estimates for middle school teachers are larger than for elementary school teachers, with the former typically varying in the range of 20-40% of a standard deviation gain in the academic

achievement of students. A second factor is the student type. Aaronson et al. (2007) and Hanushek (1992) reveal that high-quality teachers are particularly influential for lower ability students.

3) Studies relating the students' academic performance to specific teacher characteristics show that a moderate amount of variance in performance is associated with those characteristics (Rivkin et al., 2005). In these studies, the main variables considered are teacher training and certification (having passed some type of test to work as a teacher), teacher knowledge and skills, years of work experience, wages and gender.[1] In general, the empirical evidence, although mainly obtained from the United States, emphasizes the positive impact of the teachers' level of knowledge and their teaching experience (although primarily only that of the initial years). The evidence also highlights the positive effect (although not in all studies) of wage incentives related to teacher and/or student performance and, to a lesser extent, wage levels. Obtaining academic qualifications higher than that of a university degree does not seem to affect teacher quality. Likewise, the gender of the teacher, on the whole, has no impact on quality.

Below we summarize the empirical evidence for the variables that we address in our empirical analysis, namely teachers' education and training, work experience and gender.

---

[1] Most studies do not directly relate student performance with teacher characteristics, but, as we do in this study (as explained in section 4), they analyse which teacher characteristics affect their quality (identified in analyses adopting the first approach), which does improve student outcomes.

In terms of teacher training, having a master's degree or similar does not, in most studies, affect student performance (for a review of US studies, see Calero and Escardíbul 2015). Some positive evidence, however, is recorded in the case of the evaluation of mathematics in several countries participating in PISA (see Fuchs and Woessmann, 2007). Additionally, some studies conducted in the US report the positive effect on the acquisition of competencies in mathematics of a teacher holding a qualification with considerable content in that subject (undergraduate math degree or specific courses) (Calero and Escardíbul 2015).

As Leigh (2010), Metzler and Woessmann (2010) and Calero and Escardíbul (2015) show, most studies indicate that the years of experience a teacher has matter, having a positive effect on student performance. However, in most cases this effect is not linear in time, so that the improvement attributable to experience seems to be highly concentrated at the beginning of a teacher's career (no later than 3-5 years). As with training, most of these studies have been conducted in the United States.

As for teacher gender, the debate in primary education has been whether it is better for students to have a male or a female teacher, and even whether it is better for the teacher to be of the same gender as the student. Indeed, policies have been implemented in several English-speaking countries to increase the participation of men, currently in a minority, among teaching staff (Carrington et al 2007, 2008). However, there is no conclusive empirical evidence to show that male teachers are better than female teachers; moreover, in studies in which the gender variable is significant, the results suggest students benefit from having a female teacher (for a review of such studies, see Escardíbul and Mora 2013).

3. Empirical analysis: data

The data for the empirical study conducted here are drawn from the Progress in International Reading Literacy Study (PIRLS), a test organized by the International Association for the Evaluation of Educational Achievement (IEA). PIRLS analyses the reading comprehension of students in the fourth grade of primary education. The 2011 assessment is the fourth study of the learning to read process conducted by the IEA (although only three bear the same name and adhere to its current format: 2001, 2006 and 2011). In PIRLS-2011, a total of 48 countries participated. In the case of Spain, Andalusia and the Canary Islands were the only regions (Autonomous Communities) to provide data, but only Andalusia provided a sufficiently large sample to have its results included in the international report (see INEE 2013).

The structure of the PIRLS database is complex, both as regards methods of student selection and the way individual results are obtained. In the case of selection, the participating schools are first selected and, then, one or more classes within each school are chosen at random. In schools with few students in each class, students are selected from several classes and included as they if represented one class. In Spain, 312 schools were selected and 403 teachers and 8,580 students participated.

As for the students' results, responding to the full test would require a minimum of six hours. Thus, the objective of PIRLS is not to evaluate each student individually but rather a country's cohort of students. Therefore, each student only completes a part of the test. Consequently, the PIRLS assessment does not provide a direct test score for the students; rather, it generates a range of plausible values for each student, calculated on the basis of their answers and the difficulty of the questions. The responses are then adjusted or calibrated (e.g. non-responses) to improve the quality of results. In short, a

range of plausible values is provided to reduce the measurement error as well as the inference bias due to the impossibility of measuring the students' ability using a test with a limited number of questions. The overall average result of the countries participating in the PIRLS assessments is 500 points, with a standard deviation of 100. In the 2011 assessment, the average score in Spain was 513 points, with a standard deviation of 68. This score positions Spain among the group of countries that achieves an average of over 500 points, although below the average of participants from the European Union (534) and the OECD (538). Spain recorded 28 points fewer than Italy and Portugal, similar countries in many respects, while with respect to England, Northern Ireland and United States, the distance was greater (more than 39 points). There was also a difference, albeit non-significant of only 7 points, with respect to France. Among the European countries, Spain was ranked above Norway, Belgium (French-speaking community), Romania and Malta (IEA 2012).

4. Empirical analysis: methodology

The empirical study is carried out in three stages. First, in a regression that estimates the determinants of student outcomes in the PIRLS-2011 assessment, we estimate teacher fixed effects, considered as an approximation of teacher quality in terms of its impact on the acquisition of student reading skills. Second, this estimation is improved by controlling for the school-level factors that the teacher fixed effects might have incorporated. These factors include such variables as the level of education attained by the families of the students attending the school. Finally, we calculate which teacher characteristics affect fixed effect or teacher quality estimations. By using this method, we seek to answer the questions raised in the introduction, that is, to determine whether

teacher quality affects the students' skills acquisition and which observed teacher characteristics affect the level of teacher quality.

As mentioned in the previous section, the first phase in the empirical analysis relates the students' results on the PIRLS-2011 assessment with their personal and family characteristics and, likewise, it incorporates teacher fixed effects. These fixed effects are captured by using dummy variables for each teacher, so that the model identifies whether the fact that the students have been taught by a specific teacher (and not another) is relevant in explaining the students' acquisition of reading skills. The model to be estimated is the following:

$$Y_{its} = \beta_0 + \beta_{1k} X_{kits} + \Phi_t + \varepsilon_{its} \qquad \varepsilon \sim N(0,\sigma^2) \qquad (1)$$

where Y_{its} is the reading comprehension score of student "i", taught by teacher "t", at school "s"; X_{kits} is a vector of "k" factors associated to student "i", taught by teacher "t", at school "s"; and, Φ_t captures the fixed effect generated by each of the teachers. Finally, ε_{its} is the error term and the estimated parameters are denoted by β.

The empirical study involves a linear regression analysis. However, given the existence of five plausible values (as indicators of the students' scores on the PIRLS assessment) and the method applied in selecting the students (first the schools are selected, then the students), the regression used needs to take into account these five possible values and the standard errors need to be estimated at the level of the school, i.e., the level at which the PIRLS sample is initially selected. In our case, the fixed effects estimations are carried out with the STATA program. All the estimations allow

robust results to be obtained. Similarly, the presence of multicollinearity is analysed and the missing cases are imputed using regression methods (as suggested in OECD 2008).[2]

In principle, the fixed effects estimated for each teacher can be considered as approximating the effect each teacher has on his or her students' acquisition of reading skills. However, these fixed effects may include effects related to the characteristics of the schools in which the teachers work (school variables are not included in equation (1) as in most schools only one teacher is included in the sample). Therefore, the characteristics related to the school (W_{st} in equation (2)) are removed from the estimated teacher parameters in (1), giving rise to a new parameter that approximates the concept of teacher effectiveness or quality more closely, by referring to the fixed effect of the teacher on student performance while controlling for school characteristics, as shown in equation (2):

$$\upsilon_t = \Phi_t - \beta_t W_{st} \qquad (2)$$

Finally, the fixed effects calculated in (1) and in (2) are related to the teacher characteristics that can be extracted from the information collected in the PIRLS survey. In this way, we can verify whether the teachers' observed characteristics explain the effect the teachers (or the quality of the teachers) have on the students' results, as shown in equations (3) and (4). In equation (3), the dependent variable is the fixed effect directly calculated in equation (1), while in equation (4), the dependent variable is derived from the fixed effects estimated in equation (2). In this way, we can verify whether we obtain notable differences in the results by including teacher characteristics

[2] This imputation is not performed for specific teacher characteristics under analysis, nor for some particular variables (such as the gender of the students or their socio-economic and cultural status).

with and without the variables related to the school-level effects (such as the educational level of the families with children at the school). In both equations, M_t includes the different teacher characteristics:

$$\Phi_t = \alpha_t + \gamma_t M_t + \omega_t \qquad \omega_t \sim N(0,\sigma^2) \qquad (3)$$

$$\upsilon_t = \alpha_t + \gamma_t M_t + \omega_t \qquad \omega_t \sim N(0,\sigma^2) \qquad (4)$$

Table 1 shows that, on average, students in private schools score slightly higher than those in public schools (18-point difference or 3.6% higher). If we consider the schools' economic and socio-cultural resources, the difference is somewhat higher, with students at schools with a higher resource level obtaining 31 points more (6.3%) than those who study in schools with a lower resource level.

[table 1 about here]

The empirical analysis undertaken here broadly follows the proposals of Slater et al (2012) and Aaronson et al (2007). These two studies are, however, somewhat more complete as they use panel data and evaluate more than one skill, thus enriching their analyses by including student fixed effects. Likewise, both analyses consider the students' prior levels of attainment. Consequently, we expect our estimations to obtain better values related to teacher quality than those reported in these studies, given that they incorporate the effect of other variables for which they were able to control.

5. Descriptive variables

The empirical analysis considers variables at the student-level (personal and family characteristics), as well as variables at the teacher- and school-level. Here, we do not

describe all the variables included in the model, only those related to the teacher characteristics, which are the specific object under study. Table A1 in the Appendix includes the main descriptors of the variables used in the regression analyses related to the students, teachers and schools.

Table 1 shows the main statistics for the given teacher characteristics for the whole sample, elevated to the teacher-level, as well as for some subsamples. The number of years working as a teacher appears as a continuous variable, despite the fact that in some regression analyses various categories are considered. As shown in Table 2, the average age of teachers and their average number of years of teaching experience are slightly higher in private than in public schools. This difference is more marked in schools with a medium to high economic/socio-cultural status than it is in those with a medium to low status. As expected, a large majority of teachers are female (80.3% of the total). However, there is a marked difference between schools of different ownership model, so that in private schools the percentage of women is 10.8 points lower than in public schools. Yet, there is virtually no difference between schools of different socio-economic status. Differences by type of school (in both classifications) are notable with regard to the presence of teachers with a master's or PhD degree – private schools and schools with a medium to high status recording higher rates. As for the number of hours of training seminars received on reading comprehension, a difference of around 2-3 percentage points is observed in favour of public schools and schools with a medium to low resource status. Finally, in the case of the variable that considers whether the teaching of reading has been important in the teachers' university studies, a slightly higher percentage was observed for teachers in private schools and in those with a medium to low resource status.

[table 2 about here]

6. Model estimation results

This section describes the results of the model estimation. Following the series of equations presented in the methodological section, first, we present the regression analysis for the estimation of equation (1), in which the effect of each teacher on the students' reading comprehension is considered (fixed effects equation). Second, we removed the school-level factors from the estimated teacher fixed effects (equation (2)). Finally, we analysed which teacher characteristics affect the fixed effect estimates (taking into account the parameters estimated initially and the residual variance that does not incorporate the school-level effects).

The results of the estimation of equation (1) are presented in Table A2 of the Appendix. As is usual in studies of this type, the variables related to the personal and family characteristics of the students are highly significant. In our analysis a positive effect is found, in descending order, in relation to the fact the student was born in the first three quarters of the year (vs. the fourth), their level of writing competence prior to beginning primary education, as well as the socio-economic and cultural status of the student's home. The student's age has a negative impact as does the fact that the student has been the victim of bullying.

Table 3 summarizes the results of the model estimation (equation 1, "gross" fixed effects). The first column shows the mean results obtained. Three types of analysis are considered. The first row reports the variability in the teacher fixed effects (standard deviation), which allows us to know how the results would vary if we increased teacher quality by one standard deviation. The second shows the interquartile range (P75–P25 or Q1–Q3), which captures the variability in the results for students who have had a

teacher located in the 75th percentile of fixed effects and those who had a teacher located in the 25th percentile of fixed effects. The third row extends the previous analysis to consider the most extreme cases, expanding the range to P95–P5. In addition to the estimated coefficients, the standard error is indicated in parentheses, which allows us to verify the significance of all the parameters estimated. All the models are significant (according to the F-Test results). In the second column, the previous results are compared with the average plausible value, to determine the previous impact on the results of the sample under consideration. This column is relevant for comparing the results of the first column with different subsamples. The last column calculates the effects that appear in the first column in terms of the standard deviation of the dependent variable (i.e. the outcomes of the assessment). Note that the regression model estimated to achieve the results presented has an R^2 of 0.357.

[table 3 about here]

The first column shows that increasing teacher quality by one standard deviation would increase the students' scores by 25.29 points. In relative terms, this increase represents a rise of 4.9% in the results obtained (column 2) or 40% of the standard deviation of the results (column 3). If we consider the differences between quartiles, the difference (P75-P25) indicates that the gain for a student from having a "better" teacher (75th percentile) vs. having a "worse" teacher (25th percentile) is 30.1 points. In the extreme range, the difference between having a teacher from the 95th or from the 5th percentile represents a gain for students of 85.4 points.

As indicated in the methodology section, the estimated fixed effects (shown in Table 3) include the effects of each school. For this reason, we also calculate the "net" fixed effect or the effects free of the school-level characteristics, as shown in equation

(2). The results are presented in Table 4. As expected, the effect of teacher quality on students' skills acquisition is lower than that presented in the previous table, albeit only slightly. By way of example, increasing teacher quality by one standard deviation increases the PIRLS score by 22.7 points (2.55 points less than in the previous analysis). In relative terms, the effect of a one standard deviation increase falls to 4.39% compared to 4.88%. In the regression analysis the new R^2 is 0.245.

[table 4 about here]

How can the results obtained be compared with those from the literature? We saw in section 2 how, in different studies, applied to Primary Education, increasing teacher "quality" by one standard deviation translated into increases of approximately a 10-20% of a standard deviation of academic achievement. These results reach 27.2% in the study by Slater et al. (2012) applied to the United Kingdom. The equivalent results, for analyses applied to middle school, are higher, ranging between 20 and 40% (see Lockwood and McCaffrey, 2009). Our analysis indicates that increasing the standard deviation of teacher fixed effects by one unit results in a 35.93% increase in the standard deviation of the reading skill results. This result is quite higher than those reported in the literature for Primary Education and lies in the range of the results corresponding to middle school. This difference may be attributable to the fact that the methodology we have applied is not fully comparable with that used in the reference studies in the literature, insofar as our analysis does not incorporate the students' prior attainment, which means that we cannot identify, as in other studies, the value added by the teachers.

A final analysis is designed to analyze if the teachers' observed characteristics affect their quality. To do so, and as specified in equation (3), teacher variables are

included to explain the estimated fixed effects. Likewise, as equation (4) shows, this analysis is repeated with the fixed effects that come from the residual variance obtained in equation (2), to observe if there are significant differences and as an indicator of robustness.

In our study, the variables referring to the teachers include their age, years of experience as teachers, gender, certification obtained and area of specialization, as well as hours of training (over the last two years). Likewise, the regression model includes other variables, information about which comes from the questionnaire that the teachers complete in the PIRLS assessment. This includes their degree of professional satisfaction, the use they make of libraries/reading corners and computers in their classroom, their use of practical examples related to the students' daily lives when providing explanations, the streaming of students by ability in differentiated classrooms, the degree of safety/orderly behaviour in the school and the number of students in the classroom.

Table 5 shows the results of the estimation of the fixed effect models, in the first column in "gross" terms (incorporating the effect of the school-level variables) and in the second column in "net" terms (controlling for the effect of the school-level variables). If we focus on the results in "net" terms, we can see how the variables related to the age of the teachers and their being in possession of a master's degree or PhD do not affect teacher quality. However, the number of hours spent in training seminars dealing with the teaching of reading (in the past 2 years) and the fact that the teaching of reading was granted a certain importance in the teachers' studies do have an effect (positive coefficient). Being a female teacher and having little work experience (not exceeding 5 years as a teacher, a category that contrasts with the reference, corresponding to 30 years or more of experience) present a negative coefficient. Note

that the significance of the results obtained is low, given that in some cases the variables are statistically significant at the 5% or at the 10% levels, but in no case are they significant at the 1% level.

[table 5 about here]

In short, only some teacher characteristics among those analysed affect teacher quality. We can conclude that being male, having received continuous training and having specialized studies (related to reading comprehension), as well as experience as a teacher, have a positive but moderate effect. The significance in the model of the aforementioned variables is low, so we must assume that teacher quality is related to variables not observed in our study and unrelated to the more "visible" characteristics of teachers.

Our findings generally coincide (with the exception of the effect of teacher gender) with the results of other similar studies. In Section 2 it can be seen that most studies that use a similar methodology (applied mostly to English-speaking countries) show that only a moderate amount of variance in performance is associated with observable characteristics of teachers. Our study, as we have seen, points in the same direction. In the same way, the empirical evidence described in the above literature underlines the positive, though moderate, effect, of the specialised university training of the teachers, their experience, and their participation in continuous training. The estimation of our model gives results which are in accordance with this prior evidence.

However, the results of this study differ from the usual results in the literature referring to English-speaking countries as regards the effect of one specific variable. Most of the studies do not find significant differences in the effect of male and female teachers and, when a significant difference is found, it is in favour of female teachers

(see Escardíbul and Mora 2013). The significant effect in favour of male teachers that appears in our estimation is, therefore, an unexpected result and one which should be explored in further studies, in order to verify whether there is any relation with any specific feature of the Spanish education system.

7. In which schools do the "best" and "worst" teachers work?

In order to determine where teachers work according to their level of "quality", we ordered the sample in descending order in terms of their estimated fixed effects, where the "best" teachers are those whose fixed effects are situated in the top half of the scores obtained. Table 6 shows that, among the 199 "best" teachers (half the estimated "net" fixed effects, i.e. the school-level effects having been removed), 135 (67.8%) work in public schools and 64 (32.2%) in private schools. However, we need to bear in mind that 71.1% of the teachers in the sample work in public schools and 28.9% in private schools, which means the "best" teachers are overrepresented in private schools. This overrepresentation is reflected in the ratio calculated in column (4) between the number of "best" teachers (column 1) and the number that would exist if they were neither overrepresented nor underrepresented, i.e. the number that would exist if the "best" teachers were distributed homogeneously between the two types of school (column 3). This ratio is 0.97 at public schools and 1.07 at private schools. Thus, we find more "good" teachers than expected in private schools and fewer in public schools.

As for the level of socio-economic and cultural resources of the families with children at the schools, column (4) shows that there are more "good" teachers than expected in schools with higher levels of economic and socio-cultural resources

(coefficient = 1.14) than in schools with lower levels (0.84). Likewise, the distance in the provision of "better" teachers by type of school is somewhat lower in terms of the school ownership model than in that of their endowment of socio-economic and cultural resources.

[table 6 about here]

8. Conclusions

This study has sought to provide answers to three research questions, namely: i) Do teachers matter in the skills acquisition of primary school students? ii) What factors determine the quality of teachers? iii) In which schools do the "best" and "worst" teachers work? The methodological strategy adopted involved the estimation of a production function, by means of a multiple regression model in which we introduced teachers' fixed effects. The data used in the estimation are drawn from the PIRLS-2011 assessment.

The results allow us to provide the following answer to the first question: teacher quality (approximated using the model's fixed effects) is important in terms of its effects on the students' acquisition of reading skills. The gain for a student taught by a high-quality teacher (in the upper 75th percentile) is 29.77 points on the PIRLS scale with respect to a student taught by a low quality teacher (in the lower 25th percentile). If we expand the range of the comparison, the difference associated with a teacher in the 95th percentile with respect to a teacher in the 5th percentile is of 79.11 points.

These results point to a markedly higher effect than those reported in the literature based on research carried out in Primary Education in English speaking

countries. However, the methodology used is not completely comparable to that of those studies, since our model does not incorporate the students' prior attainment scores.

In response to the second research question, the study has been able to identify certain teacher characteristics that affect their quality. Being a man and having received continuous training and having specialized in the study of reading comprehension, in addition to the number of years of teaching experience, are factors that are positively associated with teacher quality. Our result concerning the effect of gender does not coincide with that obtained in studies conducted in English-speaking countries.

We are also able to offer an answer to our third research question (In which schools do the "best" and "worst" teachers work?). "Good" teachers are overrepresented in private schools and in schools in which families have higher levels of economic and socio-cultural resources. As such, the distribution of resources has a regressive effect on the students' outcomes in another example of the "Matthew Effect" (Merton 1968), insofar as the best resources are made available to the students who already have above average resources.

The results obtained throughout this article have to be interpreted in the light of the methodological limitations that prevent the establishment of causal relations. The type of information with which we work does not allow us to establish such relations, as is the case of information derived from experiments or quasi-experimental situations. For example, we need to take into consideration the possibility that students were selected for assessment by the teachers. If such selection existed, the association found between teacher quality and the results of their students could be partially due to an endogenous relationship between the prior characteristics of the students and the

characteristics of the teacher. This endogenous relationship does not seem to occur with any frequency in primary education in Spain, although with the data available to us the possibility cannot be ruled out.

The results presented in this article have clear implications for education policy. Teachers matter and their current distribution, at least in the Spanish case, has regressive effects (benefitting most those families with most resources at the outset). "Better" teachers are more likely to educate students who have better economic and socio-cultural resources, while the "worst" teachers are most likely to be found in schools where students have the worst resources at the outset. Measures to reverse this situation would have a clearly progressive effect in terms of equality. These measures might be of at least two types. On the one hand, they could intensify actions that seek to improve the quality of teachers in schools where the students come from families with fewer resources. On the other, they could promote the presence of "better" teachers in these schools (see, in this regard, Bonal (ed.) 2012, and OECD 2012).

Acknowledgements: This study was funded by the Spanish Ministry of Economy and Competitiveness through the project EDU2016-76414-R.

References

Aaronson, D., L. Barrow and W. Sander. 2007. "Teachers and student achievement in the Chicago Public High Schools". *Journal of Labor Economics*, 25 (1): 95–135.

Bonal, X. (dir.) 2012. *Municipis contra la segregació escolar. Sis experiències de política educativa local.* Barcelona: Fundació Jaume Bofill.

Calero, J. and J. O. Escardíbul. 2015. *Recursos escolares y resultados de la educación. Reflexiones sobre el sistema educativo español.* Madrid: Fundación Ramón Areces-Fundación Europea Sociedad y Educación, pp. 111-159.

Canales, A. and L. Maldonado. 2018. “Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics”. *International Journal of Educational Development,* 60 (C): 33-50.

Carrington, B., B. Francis, M. Hutchings, C. Skelton, B. Read, and I. Hall. 2007. “Does the gender of a teacher really matter? Seven-to eight-year-olds’ accounts of their interactions with their teachers”. *Educational Studies*, 334: 397-413.

Carrington, P. T. and C. Merrell. 2008. “Role models, school improvement and the ‘gender gap’: do men bring out the best in boys and women the best in girls?” *British Educational Research Journal*, 343: 315-327.

Chu, J. H., P. Loyalka, J. Chu, Q. Qu,Y. Shi and L. Guirong. 2015. “The impact of teacher credentials on student achievement in China”. *China Economic Review,* 36 (C): 14-24.

Coleman, J. S., E. Q. Campbell, C. J. Hobson., J. Mcpartlet, A. M. Mood, F. D. Weinfeld and R. L. York. 1966. *Equality of Educational Opportunity.* Washington, D.C.: Government Printing Office.

Escardíbul, J. O. and T. Mora. 2013. “Teacher gender and student performance in mathematics. Evidence from Catalonia (Spain)”. *Journal of Education and Training Studies,* 1 (1): 39-46.

Fuchs, T., and L. Woessmann. 2007. “What accounts for international differences in student performance? A reexamination using PISA data”. *Empirical Economics,* 32 (2-3): 433-462.

Hanushek, E. A. 2011. “The economic value of higher teacher quality”. *Economics of Education Review,* 30: 466-479.

Hanushek, E. A. 1992. “The Trade-Off Between Child Quantity and Quality”. *Journal of Political Economy,* 100 (1): 84-117.

Hanushek, E. A., J. Kain, D. O’Brien and S. Rivkin. 2005. “The market for teacher quality”. *NBER Working Paper 11154.* Cambridge, Mass.: National Bureau of Economic Research.

Hanushek, E. A. and S. G. Rivkin. 2012. “The Distribution of Teacher Quality and Implications for Policy”. *Annual Review of Economics,* 4 (1):131-157.

IEA. 2012. *PIRLS 2011. International results in Reading.* Chestnut Hill, MA.: International Association for the Evaluation of Educational Achievement (IEA).

INEE. 2013. *PIRLS - TIMSS 2011. Estudio Internacional de progreso en comprensión lectora, matemáticas y ciencias. Volumen I.* Madrid: Ministerio de Educación, Cultura y Deporte. Instituto Nacional de Evaluación Educativa.

Jacob, B. A. and S. D. Levitt 2003. "Rotten apples: An investigation of the prevalence and predictors of teacher cheating". *Quarterly Journal of Economics,* 118 (3): 843–77.

Kane, T. J.; J. E Rockoff and D. O. Staiger (2008). "What does certification tell us about teacher effectiveness?," *Economics of Education Review*, 27 (6): 615–631.

Koedel, C., Mihaly, K. and Rockoff, J. E. 2015. "Value-added modeling: a review". *Economics of Education Review*, 47 (C): 180-195.

Leigh, A. 2010. "Estimating teacher effectiveness from two-year changes in students' test scores". *Economics of Education Review*, 29 (3): 480–488.

Lockwood, J. R. and D. F. McCaffrey. 2009. "Exploring Student-Teacher Interactions in Longitudinal Achievement Data" *Education Finance and Policy*, 4 (4): 439-467.

Merton, R. K. 1968. "The Matthew Effect in Science". *Science,* 159 (3810): 56-63.

Metzler, J. and L Woessmann. 2010. "The impact of teacher subject knowledge on student achievement: evidence from within-teacher within-student variation". *IZA Discussion Paper Series* 4999. Bonn: The Institute for the Study of Labor.

OECD. 2012. *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, OECD Publishing. http://dx.doi.org/10.1787/9789264130852-en.

OECD. 2008. *Handbook on constructing composite indicators. Methodology and user guide.* Paris: Organisation for Economic Co-Operation and Development.

Rivkin, S., E. Hanushek and J. Kain. 2005. "Teachers, schools, and academic achievement". *Econometrica*, 73: 417–458.

Rockoff, J. 2004. "The impact of individual teachers on student achievement: Evidence from panel data". *American Economic Review*, 94 (2): 247–252.

Slater, H., N. M. Davies and S. Burgess. 2012. "Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England". *Oxford Bulletin of Economics and Statistics*, 74 (5): 629-645.

APPENDIX

Table A1. Descriptive statistics of the explanatory variables

Students				
Variables	**Minimum**	**Maximum**	**Mean**	**Standard deviation**
Girl	0	1	0.495	0.499
Age	6.500	13.250	9.847	0.447
Born in the first quarter	0	1	0.240	0.427
Born in the second quarter	0	1	0.265	0.441
Born in the third quarter	0	1	0.238	0.426
Born in the fourth quarter	0	1	0.258	0.437
Home language corresponds to language of test	0	1	0.687	0.464
Student has been victim of bullying	0	1	0.569	0.495
More than three years of infant education	0	1	0.676	0.468
Age of student when starting primary school no greater than 5	0	1	0.468	0.499
Child's ability to write when starting primary school: very good	0	1	0.425	0.494
Child's ability to write when starting primary school: moderately good	0	1	0.390	0.488
Child's ability to write when starting primary school: not very good or none existent	0	1	0.185	0.388
Index of socio-economic and cultural status	3.610	15.140	10.292	1.734
Teachers				
Variables	**Minimum**	**Maximum**	**Mean**	**Standard deviation**
Teaching experience (years)	1	45	19.945	11.502
Female teacher	0	1	0.803	0.398
Age (1)	23	63	45.590	10.326
Holder of a master's or PhD	0	1	0.014	0.117
Teaching/pedagogy of reading is important in education or training	0	1	0.539	0.498
Number of hours spent in training seminars dealing with teaching of reading (in past 2 years)	0	53	16.508	19.982
Degree of satisfaction with	0	1	0.700	0.458

professional career				
Number of students per class	2	46	22.668	6.309
Students are streamed by ability: often or always	0	1	0.129	0.335
Students are streamed by ability: sometimes	0	1	0.492	0.499
Students are streamed by ability: never	0	1	0.379	0.485
Use of computers in the classroom	0	1	0.564	0.496
Use of libraries/reading corners in the classroom	0	1	0.373	0.484
Lessons are often related to the students' daily lives	0	1	0.806	0.395
The school is in a safe neighborhood and the students behave in an orderly manner	0	1	0.457	0.498
Schools				
Number of students in the school	14	2.381.0	585.592	443.419
% of students having the language of the test as their native language	0	1	0.598	0.490
Immediate area in which school is located: urban	0	1	0.311	0.463
Immediate area in which school is located: suburban	0	1	0.104	0.305
Immediate area in which school is located: medium-sized city/large town	0	1	0.449	0.497
Immediate area in which school is located: small town or rural village	0	1	0.137	0.343
Shortage of teaching resources: Not at all	0	1	0.707	0.455
Shortage of teaching resources: Some	0	1	0.156	0.363
Shortage of teaching resources: A lot	0	1	0.137	0.343
Teachers' degree of success in implementing the school's curriculum	0	1	0.695	0.460
Student achievement used to evaluate teachers	0	1	0.780	0.414
School's social make-up: high level of resources	0	1	0.513	0.500
School's social make-up: intermediate level of resources	0	1	0.305	0.460

School's social make-up: low level of resources	0	1	0.182	0.386
Public school	0	1	0.657	0.475

(1) Continuous variable based on intervals.

Note: Both the student and school descriptive statistics are high overall in the overall population at the student-level; the teachers' descriptive statistics are high at the teacher-level.

Table A2. Determinants of student outcomes

Variables at student-level	**Coefficient (E.S.)**
Girl	1.900 (2,205)
Age	-20.910*** (5.056)
Born in the first quarter	28.060*** (4.466)
Born in the second quarter	20.850*** (3.725)
Born in the third quarter *(Base category: Born in the fourth quarter)*	8.210*** (3.107)
Home language corresponds to language of test	3.854 (3.240)
Student has been victim of bullying	-7.212*** (2.111)
More than three years of infant education	-0.405 (2.054)
Age of student when starting primary school	-2.920 (2.226)
Child's ability to write when starting primary school: very good	32.410*** (3.615)
Child's ability to write when starting primary school: moderately good *(Child's ability to write when starting primary school: not very good or none existent)*	12.630*** (2.755)
Index of socio-economic and cultural status	9.324*** (0.806)
Includes teachers' fixed effects	Yes
Includes constant	Yes
N	7.512
R^2	0.357

*** Significant at 1%.

Table 1. Students' average scores: total and by type of school

Simple	Results
Total	513
Public schools	507
Private schools	525
Schools with a medium to high economic and socio-cultural resource index	525
Schools with a medium to low economic and socio-cultural resource index	494

Table 2. Teacher characteristics. Mean values.

Teacher variables	All schools	Public schools	Private schools	Medium/high status schools	Medium/low status schools
Age (years)	45.6	45.0	47.0	48.2	41.8
Female teachers (%)	80.3	83.5	72.7	80.3	80.0
Teaching experience (years)	19.9	19.2	21.9	22.6	15.9
Holder of master's or PhD (%)	1.4	0.3	4.0	2.1	0.4
University studies with specific training in teaching of reading (%)	53.9	52.8	56.7	53.8	54.1
Number of hours spent in training seminars dealing with reading comprehension (in past 2 years).	16.5	17.5	14.2	15.7	17.8

Table 3. “Gross” teacher fixed effects and variability.

Variability	**(1) Mean**	**(2) Over relative mean (%)**	**(3) Over standard deviation (%)**
Standard deviation of the teacher fixed effects	25.29 (0.90)	4.88	40.00
Interquartile range (P75 – P25)	30.07 (1.29)	5.80	47.53
Extreme range (P95 – P5)	85.44 (2.37)	16.48	135.04

Table 4. “Net” teacher fixed effects and variability.

Variability	**Mean**	**Over relative mean (%)**	**Over standard deviation (%)**
Standard deviation of the teacher fixed effects	22.74 (0.83)	4.39	35.93
Interquartile range (P75 – P25)	29.77 (1.68)	5.74	47.05
Extreme range (P95 – P5)	79.11 (3.08)	15.26	125.04

Table 5. Teacher characteristics affecting teacher quality

	Coefficient (Standard error)	
Variables	**Gross fixed effects of dependent variable (with school-level effects) Equation (1)**	**Net fixed effects of dependent variable (without school-level effects) Equation (2)**
Age	-0.076 (0.262)	0.029 (0.250)
Years of teaching experience: 1-5	-21.295*** (8.065)	-12.835* (7.641)
Years of teaching experience: 6-10	-12.499* (7.488)	-9.894 (7.038)
Years of teaching experience: 11-20	-17.503*** (5.546)	-5.851 (5.425)
Years of teaching experience: 21-30 *(Base category: 30 years or more)*	-8.780* (4.545)	-4.536 (4.261)
Female teacher	-7.968** (3.771)	-7.682** (3.437)
Holder of a master's or PhD *(Base category: degree or similar and below)*	13.194 (13.036)	1.127 (11.663)
University studies with specific training in teaching of reading	7.201** (3.057)	5.367* (2.824)
Number of hours spent in training seminars dealing with reading comprehension (in past 2 years)	0.233*** (0.074)	0.181** (0.070)
Constant	Yes	Yes
Other teacher variables included	Yes	Yes
R^2	0.215	0.151
N	332	332

*** Significant at 1%. ** Significant at 5%. * Significant at 10%.

Table 6. Teacher quality by school type

	(1)	(2)	(3)	(4)
Sample	**Teachers among the best 50%**	**Teachers by school type (%)**	**Potential teachers among the best 50%**	**Teachers among the best 50% over total number of potential teachers**
Public school	130	71.1	134	0.97
Private school	58	28.9	54	1.07
Schools with medium to high economic and socio-cultural resources	117	55.0	103	1.14
Schools with medium to low economic and socio-cultural resources	71	45.0	85	0.84

Notes:

- the percentage of teachers in each type of school as a function of the socio-economic and cultural level is not evenly distributed due to the greater number of cases lost in the sample of schools with a lower resource levels.
- the table shows the fixed effects calculated using equation (2) (i.e. “net” effects). These results vary only very slightly from those obtained using equation (1) (i.e. “gross” effects).