Do teachers' effectiveness and efficiency go hand in hand? Evidence from fourth grade Spanish schools

Lopez-Agudo, Luis Alejandro [1,2], Marcenaro-Gutierrez, Oscar David [1*]

[1]*Departamento de Economía Aplicada (Estadística y Econometría, 15)*

[2]*Programa de Doctorado en Economía y Empresa, Universidad de Malaga*

Universidad de Malaga, C/ Ejido, 6, 29071, Malaga, Spain

Abstract

Effectiveness and efficiency are desirable properties for education systems. Because of that, the aim of this paper is to determine which factors –among a wide set provided by both TIMSS 2011 and PIRLS 2011– determine teacher effectiveness and efficiency in the Spanish education system. Concretely, as teachers are in direct contact with students and are responsible for transmitting knowledge and skills to them, we intend to discern which teachers' characteristics and teaching methods contribute the most to improve their effectiveness and efficiency in enhancing students' achievement. Besides, we aim to determine whether those teachers who are effective in the development of the teaching-learning process are also efficient and vice versa in primary education, in order to provide a trustful insight in the relation of these characteristics to develop precise education policies.

Keywords: teacher, effectiveness, efficiency, primary education

JEL codes: I21

---

[*]Corresponding author e-mail: odmarcenaro@uma.es; Tel.: +34 95 213 7003.

Coauthor e-mail: luisalej@gmail.com; Tel.: +34 95 213 7003.

We acknowledge the research funding provided by the "Proyecto Puente de la Universidad de Malaga", whose main researcher is Oscar David Marcenaro-Gutierrez, and also SEJ-532 research group. Luis Alejandro Lopez-Agudo also acknowledges the research training provided by the "Programa de Doctorado en Economía y Empresa de la Universidad de Malaga"

1. Introduction

Effectiveness and efficiency measurement constitute concepts commonly related to the world of business and productive processes. However, these terms have been increasingly applied to other fields as could be education and the different levels which compound it –schools, classes, teachers, etc.–. Despite of their relevance, they are usually confused in the economics of education literature. Because of that, we will begin by providing a precise definition for both aspects applied to teachers in order to delimit their range. Beginning with effectiveness, an effective teacher could be defined as the one who has the ability to reach out to the students and make them learn. These teachers can take what students know –prior knowledge–, link it to what they do not know and then apply it to their lives in such a way that requires them to think and learn on a higher level, by using procedures which help students exert effort and learn. There are many practices included in this concept, as teachers' ability solving students' doubts, discussing the homework with them, being interested in what they teach, etc. In other words, those practices which define teachers' behavior in the classroom and the way they conduct their lessons.

Unlike, the concept of teacher efficiency basically means getting the maximum output from a quantity of inputs the teacher uses in the teaching-learning process or achieving the same output with a minimum quantity of inputs. Koopmans (1951) also provided a definition for technical efficiency, which states that an input-output vector is technically efficient if –and only if– reducing any input or increasing any output is possible only by increasing some other input or reducing some other output. Thus, teachers' efficiency can be reflected in variables which measure the way teachers use their available resources and time in order to teach their students. These variables could be the time teachers spend correcting students' homework, how they cope with a high number of students in the class, the use of the workspace and instructional materials, etc. Hence, it is the ability to get things done in an organized way, using his/her available resources and time without wasting any of them. It does not mean doing little work; it means making the work meaningful.

The main idea behind this paper is to analyze whether a teacher can be effective without being totally efficient or vice versa. In this way we intend to disentangle to what extent both concepts imply each other or not and which of these characteristics would be more desirable when asking for them to teachers. In order to do that, we will provide instruments which could perform the measurement of both characteristics, aimed to the definition of a proper system to evaluate teachers' contribution to students' achievement, so that teachers can be correctly classified into rankings.

The importance of these rankings based on efficiency and effectiveness can be found in the fact that raw rankings –or league tables–, which have been traditionally used to measure teacher or school effectiveness, do not consider many important issues as the need to account for sampling variability and other sources of error (Goldstein and Spiegelhalter, 1996; Leckie and Goldstein, 2009). A proper classification of teachers is essential to foster education system effectiveness and efficiency, due to it conditioning many facts as salaries and other incentives to promote teacher quality. As Dolton and Marcenaro (2011) concluded –from their analysis of 39 countries' panel data–, the recruitment of higher ability teachers and letting quicker salary increase will have a positive effect on students' outcomes. Other authors have also shown the positive relationship between teacher performance-related salary and higher students' achievement –and also lower dropout rates– as Loeb and Page (2000) or Lavy (2002, 2004).

Focusing on the methodology employed in this study, on the one hand, teachers' effectiveness is going to be measured by making use of multilevel regression models and, on the other hand, we estimate teachers' efficiency focusing on stochastic frontier analysis (SFA). The latter could help us to define a parametric function, which could be easily comparable with to the parametric form of the multilevel regression model, rather than other non-parametric approaches, as data envelopment analysis (DEA) or free disposal hull (FDH) methodologies. Although the definition of a proper functional form for SFA could also make problems arise, measurement errors are a major concern when dealing with students' score data, which is a field where SFA outperforms non-parametric approaches. This fact joined with the use of a moderate

set of inputs –which can be handled by SFA– make stochastic frontier analysis more suitable for the study we intend to perform.

To accomplish this research we make use of PIRLS (Progress in International Reading Literacy Study) and TIMSS (Trends in Mathematics and Science Study) 2011 for fourth grade –9/10-year-old Spanish students–. These datasets were chosen because students are more malleable at this age (Thompson-Schill *et al.*, 2009), so they would show a better reflection of teachers' procedures. Besides, primary school students normally have the same teacher for all the subjects[1]. In addition, both statistical operations have a rich set of teacher level variables gathered in their teacher questionnaire. Reading scores were selected as dependent variable in the multilevel regression models and as output in the stochastic frontier analysis. The reason for this is that most teachers of the sample are teaching reading. Due to the reduced literature which has made use of the mixed database of TIMSS and PIRLS 2011 and the virtual not existence –to the best of our knowledge– of previous literature on teacher effectiveness versus efficiency in the Spanish case, this paper will make a notable contribution to the existent literature.

Unfortunately, the nature of our data do not let us obtain test scores for more than one year for the same students –making impossible to perform a panel data analysis or value-added models, widely used in effectiveness literature by authors as Marcenaro (2002) or Lenkeit (2012b)–. In addition, neither TIMSS nor PIRLS 2011 provided figures on the number of years which students passed with the same teacher. However, in PIRLS 2006 approximately 70% of the Spanish teachers had been teaching the same class at least for two years –Spain did not participate in PIRLS 2001, so we do not have data on for this year–. Thus, as the Spanish education system has not suffered a great change in its structure or regulation which could change this trend in the way teachers are assigned to a classroom for a period of time, we could overcome the lack of students' prior achievement information by assuming that students are taught by the same teachers for many years –at least two–.

[1] This can be seen in the figures of PIRLS and TIMSS 2011: From a total of 200 teachers, 187 declared that they taught reading, 187 mathematics and 182 sciences.

To sum up, the analyses of the relationship between teachers' effectiveness and efficiency will allow us to determine which of these characteristics is more desirable, in order to propose policy instruments which could help to set optimal teacher rankings and, to some extent, it could be used to define salary policies. We intend to answer these questions: Are the students of the most effective teachers those who get better results, those of the most efficient, or both of them? Which teachers' characteristics have more impact on effectiveness and/or efficiency? The rest of the paper is organized as follows: the second section will provide a literature revision of both teacher effectiveness and efficiency measures in education system. The third section will show the methodology employed to perform this analysis, while the fourth section will deal with the data. Fifth section will show the results obtained from estimating the proposed models. We will end up –section six– with a summary of the most relevant conclusions and policy implications of this analysis.

2. Literature revision

Beginning with the revision of effectiveness in the education literature, it has provided many definitions and elements which characterize effective teachers. These teachers are said to create a task-oriented environment and devote classroom time on academic activities rather than on socializing, free time, etc., as it has been highlighted in a wide range of contexts and countries (Creemers, 1994; Scheerens and Creemers, 1996; Reynolds *et al*., 1996). They can also organize their classrooms as effective teaching environments in which academic activities are developed with brief transitions (Brophy and Good, 1986), dealing at the same time with potential bad behavior or other problems which could affect the flow of the lesson (Borich, 1996).

The contribution of effective teachers to students' achievement is so important that students learn more in classes taught or supervised by their teachers rather than doing homework or other school tasks at home, due to teachers providing the knowledge directly to the student –as opposed to other methods of self-learning– assuming an active role (Brophy and Good, 1986; Borich, 1996) and linking new knowledge to students' prior knowledge. However,

individual or group tasks are essential in effective teaching, being necessary to explain and monitor the tasks, and also to provide feedback to students. Following Beaman and Wheldall (2000), interventions as acknowledging students have been shown to decrease students' inappropriate behaviors and increase appropriate ones, to the extent that praising constitutes a positive reinforcement for students' appropriate behavior. However, this does not mean that teachers should use praising indiscriminately, because it could become meaningless.

Boonen *et al*. (2013) also analyzed the effect of teacher characteristics (background qualifications, attitudes and beliefs, and instructional practices) on student achievement in mathematics of first grade, reading and spelling for first grade in the Flemish region of Belgium. They found that teacher background had the largest effect on mathematics achievement, while instructional practices had higher effects on both reading and spelling achievement. However, the larger teacher effects were found for mathematics than for spelling and reading.

Teacher effectiveness has been frequently measured in the literature by using cross-sectional data from student and family characteristics (Nachtigall *et al*., 2008). However, when data for two years or more are available, most research propose the formulation of value-added models, to the extent that prior attainment is considered as an essential factor which conditions subsequent achievement. As Manzi *et al*. (2014) highlight, value added has become a broadly used methodology in educational research on effectiveness, which is often modeled as a random effect in a specific hierarchical linear model. However, they also showed that this approach was not valid when endogeneity is present to the extent that this would mean a correlation between the random effect in the hierarchical model and some of the covariates, what can be a generic problem when the covariates contain the prior scores –what is usual in value added modeling–

In addition, Teddlie *et al*. (2000) state that prior achievement data for primary schools could be problematic to obtain, to the extent that the information collected prior to the school entry is low in quantity, so they declare that using background characteristics will be more appropriate in order to measure effectiveness in primary schools. Lenkeit (2012a) also focused on the subject of the differences between value-added models and contextualized attainment

models –those which contain only cross-sectional data–, reaching to the conclusion that the latter are a proper alternative to value-added models when the only available data is referred to one year, because students background characteristics act as good substitutes for prior attainment. So, data from international evaluations could be used to measure effectiveness in a reliable way.

In the field of teacher effectiveness, some studies which have used the cross-sectional data from TIMSS, e.g. Zuzovsky (2009), studied the effectiveness presented by teachers from Israel by using a multi-level regression analysis. She defined a model which included teacher's formal education, teachers' subject, education in pedagogical studies, duration of pre-service education, certification and licensing status, years of experience and participation in professional development activities. They found that teachers' advanced academic training and years of experience in mathematics and science have a differential effect due to their inner characteristics. Since science is a constantly developing domain, science teachers with advanced and frequently updated education may have higher results, while mathematics is considered a basic subject, so teaching experience is more important than updated knowledge of the domain.

Focusing on teachers' efficiency, it has a tradition in the literature related to the study of the management of schools in order to assigning costs and resources. However, we intend to measure the capacity of teachers to use their available materials, teaching resources and to organize their time in order to get the best results from their students. Naper (2010) studied the effect of teacher hiring practices by 10^{th} grade school directors and educational efficiency of Norwegian school districts. They applied a school level analysis with district fixed effects and found that the efficiency was higher in schools facing excess of teacher supply and where decentralization was stronger.

Jürges and Schneider (2008) developed a stochastic frontier model in order to define a fair teacher ranking based on the efficiency of each teacher. They intended to develop a basis for performance related payment schemes in Germany, so they employed German PIRLS data for 2001 accounting for the socio-economic background of students and schools and obtained

clear differences between the ranking based on the unconditional test scores and the ranking based on the efficiency estimates. The reason for this was that less than two thirds of the ranked teachers remain in the same quality group after controlling by student and school background, meaning that rankings based on raw test scores were unfair in the sense that they did not account for relevant background information. Because of that, we will take into account students' background information, in order to proportionate a precise ranking of efficiency.

However, although there is not literature which intend to compare teacher effectiveness and efficiency, this approach has been applied to other levels of the education production function, as can be found in the field of school efficiency, where McEwan and Carnoy (2000) analyzed Chile's public and private schools and compared efficiency and effectiveness, finding that public schools are more effective, although their efficiency is similar.

Thus, we can infer from the previous literature that these concepts of effectiveness and efficiency in teaching have been approached by many studies in a separate way and they have been frequently confused because of their imprecise definition and methodologies employed for their measurement. In our study we intend to relate each teacher characteristic to its correct term (effectiveness and efficiency) in order to outline the best measurement for each one.

3. Methodology

When performing our analysis we are going to employ two different methods which have been widely used by the literature. In order to provide teacher effectiveness measures we are going to make use of multilevel regression analysis, while stochastic frontier models will be applied for teachers' efficiency. Each of these methodologies will control by the same student background variables and a general set of teacher variables, but the other variables in the specifications will change when teacher effectiveness or efficiency is studied. The dependent variable in both effectiveness and efficiency analysis are students' scores in reading –aggregated by teacher for stochastic frontier analysis–. This variable has been widely employed in the literature of education effectiveness (Zuzovsky, 2009; Lenkeit, 2012a; Braga *et al*., 2014) and efficiency (Jürges and Schneider, 2008; Naper, 2010).

3.1. Multilevel regression analysis

Raw rankings –sometimes called league tables– have been traditionally used to measure teacher or school effectiveness. However, they ignore many issues as the need to account for sampling variability and other sources of error (Goldstein and Spiegelhalter, 1996; Leckie and Goldstein, 2009). It is also needed to make a proper adjustment of raw outcomes, so the main statistical tool for this task –and also for quantifying uncertainty– is regression. However, standard regression models do not take into account the hierarchical structure which educational data usually presents.

Following works as Lenkeit (2012a), multilevel regression analysis (Hierarchical Linear Models, HLM) constitutes a proper approach in order to analyze effectiveness when cross-section data from international studies are available. This methodology has been widely used due to its capacity to account for the existence of a hierarchical organization of the data in the teaching-learning process (Raudenbush and Bryk, 2002; Goldstein, 2003; De Leeuw and Meijer, 2008; Snijders and Bosker, 2012). It is usually employed in a stepwise procedure: departing from a basic specification, groups of variables –corresponding with each of the levels studied– are incorporated to the specification in each phase. Thus, this approach does not offer a unique regression line –as happens in OLS–, but a multiple amount of them: one for each studied level. These models provide information about the average effect of the explanatory variables on the dependent one and the variation of these effects at the proposed levels.

Concretely, we will estimate the contribution to the variance (total heterogeneity) of students' characteristics in the first level and that of teachers in the second level. Then, a regression for each teacher in the second level will be defined. The estimated model can be specified in a general way by:

$$Y_{ij} = \beta_{0j} + X'_{ij}\beta + \varepsilon_{ij} \qquad (1)$$

where the matrix Y_{ij} denotes the mean of the five plausible values obtained in the reading subject by a student "i" who is taught by teacher "j"; β_{0j} is a parameter of level that reflects the

differential effect on Y_{ij} of each teacher, i.e., it is a proxy of the effectiveness of the teacher; X_{ij} is a vector of student characteristics and ε_{ij} is the idiosyncratic error term.

However, the specification of (1) assumes that the intercept of the regression is the same for all transversal units, but it is very likely that we need to control the "individual" character of each state. Besides, random effects model suggests that each teacher has a different level parameter (interception term), which it is not fixed and behaves as a random variable with an average value (α) affected by deviations represented by the random variable u_j, so β_{0j} could be defined as:

$$\beta_{0j} = \alpha + u_j \tag{2}$$

When substituting (2) in (1) we get the expression of the random effects model:

$$Y_{ij} = \alpha + \beta_1 X_{1ij} + \cdots + \beta_p X_{pij} + u_j + \varepsilon_{ij} \tag{3}$$

The random effects estimation is preferable in this case to the extent that the hypothesis of zero correlation between the disturbance term associated with the teacher (u_j) and the exogenous variables (X_{ij}) could not be rejected. The ε_{ij} is assumed independent among students and also independent of u_j. Finally, we add the effect on the scores (Y_{ij}) of the variables considered at the teacher level (Z_j) to the expression (3), reaching the following model:

$$Y_{ij} = \alpha + \beta_1 X_{1ij} + \cdots + \beta_p X_{pij} + \delta_j Z_j + u_j + \varepsilon_{ij} \tag{4}$$

which represents the multilevel model of estimation using random effects.

In order to define a ranking of teacher effectiveness, teacher-level residuals are usually used in the literature, because they can be defined as predictions of the random effects representing the effectiveness. Following Raudenbush and Willms (1995), these residuals can be obtained by two different procedures: by subtracting the expected outcome from the observed outcome or using empirical Bayes (EB) method –which has been our election–. The first method provides unbiased estimates of teacher effectiveness, while the EB method produces the so called "shrunken residuals". These residuals are biased but efficient estimates of

teacher effectiveness. Depending on the cluster size, this reduction pulls the conventional residual towards zero: the smaller its size, the greater the reduction. However, one positive fact of "shrunken residuals", apart from their efficiency, is that they avoid the accidental assignment of a teacher to the top or bottom of the ranking, although these residuals are affected by the sampling variability and other errors. Because of this problem, Goldstein and Healy (1995) suggested creating a pairwise confidence intervals in order to overcome this uncertainty. We follow this method to provide a proper ranking of top and bottom performer teachers, which is the aim we are pursuing.

Empirical Bayes estimation has been employed by authors as Arpino and Varriable (2009), who assessed the development of rankings for the second level units in hierarchical linear models by the empirical Bayes prediction. They highlight the precision of this method to rank those second level units (schools, teachers, etc.) which are top performing in effectiveness and also those with the lowest effectiveness. This methodology was also employed by Lenkeit (2012b) to analyze the effectiveness of educational systems participating in PIRLS 2001 and 2006.

3.2. Stochastic frontier analysis

As previously stated, we make use of stochastic frontier analysis in order to measure teacher efficiency, following works as e.g., Jürges and Schneider (2008). This parametric technique is based on the premise that the deviations of the production function are due to statistical noise, which makes the model of educational production function stochastic. Therefore, a stochastic factor cannot be attributed to educational production process, and should not be included in a single error term. This frontier model can show the best combination of inputs that maximize the average student performance.

Formally, the SFA models is presented as follows:

$$y_i = \alpha + x'_i\beta + z'_i\gamma + \varepsilon_i$$

$$\varepsilon_i = v_i - u_i$$

$$v_i \sim N(0, \sigma_v^2)$$

$$u_i \sim W$$

For $i = 2, \dots, N\ teachers$

Where y_i denotes the logarithm of the mean output of the students taught by the ith teacher; x_i is the input vector of students' background mean characteristics and β the vector of their technology parameters, while z_i is the input vector of teacher general variables and resources –time, materials, etc.– and γ the vector of their technology parameters. The composite error is ε_i, which consists of the sum or the difference of the disturbance u_i (representing inefficiency) and v_i, which is a normally distributed disturbance, being these two terms independent one of another and $i.i.d.$ between observations. The inefficiency term u_i can be distributed according to a Half-Normal, a truncated normal, a Gamma or as an exponential distribution, which is the one we are going use, following Meeusen and Van den Broeck (1977). Due to the large number of assumptions about the distribution of the inefficiency term for identification, it will be necessary to perform the estimation by maximum likelihood.

4. Data

Our analysis will focus on analyzing the effectiveness and efficiency of the Spanish teachers who took both two of the statistic operations offered by the International Association for the Evaluation of Educational Achievement (IEA): TIMSS and PIRLS. The purpose of TIMSS is to measure learning achievements in the areas of mathematics and science of students at the end of fourth (9-10 years) and eighth grade (13-14 years), while PIRLS is focused on reading achievement on fourth grade. TIMSS has been conducted every four years since 1995, when the series of four-year cycles began. PIRLS' periodicity is five years and it has been conducted since 2001. The design of these evaluations allows for macro-comparisons over time and among countries participating in it, but the students, teachers and schools who participate are different in each cycle. As we previously remarked, we are going to make use of TIMSS and PIRLS 2011 data on fourth grade students and teachers units for reading.

In each cycle of PIRLS and TIMSS their assessment framework is updated considering the curricula of the countries and the latest developments in the measurement of learning in reading, mathematics and science, maintaining the continuity and comparability with previous testing cycles. By 2011, PIRLS and TIMSS coincided for the first time, so that participating countries have had the opportunity to gather information from both studies for the same sample of students and thus make comparative analysis between the performances of students in the three subjects. This is the case of Spain, whose PIRLS sample was composed by a total of 8,580 students and TIMSS sample by 4,183 students of the PIRLS sample, so students who participated in both of them were chosen for our study –who were taught by a total of 200 teachers–. The use of a mixed database of TIMSS and PIRLS has been employed by studies like Foy and O'Dwyer (2013) and Grilli *et al*. (2014) in order to analyze the three subjects.

The TIMSS and PIRLS studies are composed altogether by four models of context questionnaires: students, families (only in PIRLS 2011), teachers and school. The survey questions were designed to take into account the key aspects of these contexts that may condition the competences evaluated. Thus, students are asked about family, school environment and experience in learning to read (in PIRLS), mathematics and science. Teachers are asked about their educational background, experience, resources, methods used in teaching, etc. The school questionnaire, taken by the principal or other member of the management team, asks about school characteristics, resources, educational environment, relationships with parents, etc.

In what follows, we describe the selected variables for each of the levels –students and teachers– of the teaching-learning process we are dealing with in our analysis. In order to control for student characteristics we make use of the Home Resources for Learning index

provided by PIRLS 2011[2]. The main reason for using this approach is that our analysis is focused on teachers' level, so following the literature, this variable will be a good "proxy" of students' background characteristics –which resulted significant in the analysis of Foy and O'Dwyer (2013) and Grilli *et al*. (2014), among others–. In addition, its use will let to keep a higher number of degrees of freedom. This variable will be used as student background control in both effectiveness and efficiency specifications.

When analyzing teacher level, we divide each of the available teacher variables according to their adscription to teachers' general, effectiveness or efficiency characteristics –supporting this classification on the previous literature in these subjects–. Beginning with the variables related to teacher general characteristics, we consider their sex and years of experience. Klassen and Chiu (2010) found that teacher years of experience had an effect on their effectiveness, which increased from early career until falling afterwards, denoting a nonlinear behavior. They also stated that female teachers had higher classroom stress and lower classroom management self-efficacy. However, others obtained that sex was not a factor which could condition teachers' effectiveness (Slater *et al*., 2012). These variables will be included in both effectiveness and efficiency specifications. They are often accompanied by others which account for teacher highest level of education (Jacob and Lefgren, 2008). However, in the case of our data, 184 out of 188 teachers –of the total 200 who took both TIMSS and PIRLS 2011 in Spain– had ISCED level 5a(first), and only 4 had ISCED level 5a (second), what makes irrelevant to control by this characteristic.

For the rest of the variables, when determining whether a teacher variable represent his/her effectiveness or efficiency, we have asked ourselves the following question: Does this

[2] This index is derived from three variables from parents' questionnaire –parents' education, parents' occupation and number of children's books at home– and two from children's questionnaire –number of books in the home and availability of key study supports at home: an internet connection and their own room. A figure lower than 7.3 means "Few resources", between 7.3 and 11.9 "Some resources" and higher than 11.9 "Many resources".

variable describe a method or procedure defined by the teacher when teaching his/her students to make them exert effort on learning –effectiveness– or does it represent the resources available for teachers in order to develop their lessons, as materials, classroom characteristics or time, which are not easily alterable by the teacher –efficiency–?

5. Results

We discuss, now, the main results obtained from both teacher effectiveness (Table 1) and efficiency analysis (Table 2) for the reading subject, performed for the selected variables.

Table 1. Teacher effectiveness estimates for Reading

Variables	Reading Score
Home Resources for Learning index	11.469***
	(0.635)
Female teacher (Reference: Male teacher)	-12.662**
	(5.353)
Years of experience	0.382
	(0.737)
Squared years of experience	0.002
	(0.017)
Use questioning to elicit reasons and explanations (Reference: Some lessons)	
Every or almost every lesson	20.849**
	(8.877)
About half the lessons	3.047
	(10.090)
Frequency of homework assignment:	
Two or less times a week (Reference: More than twice a week)	-9.217**
	(4.194)
Emphasis to monitor students' progress in reading:	
National or regional achievement tests (Reference: Little or no emphasis)	
Major emphasis	-17.934**
	(8.085)
Some emphasis	3.071
	(4.259)
Discuss the homework in class (Reference: Sometimes)	
Always or almost always	-4.911*
	(4.004)
Constant	391.914***
	(14.170)
Observations	3,041
Number of groups	161

Source: Authors' own calculations from PIRLS and TIMSS 2011 data.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results presented in Table 1 show the estimates for teacher effectiveness in reading. As it can be appreciated, the index of Home Resources for Learning is positive, as it could be expected, so a higher cultural level of the family increases students' achievement. Students of female teachers obtained a lower performance than males' and teachers' years of experience do not seem to affect students' achievement.

Table 2. Teacher efficiency estimates for Reading

Variables	Ln Average Reading Score
Ln Home Resources for Learning index	0.334***
	(0.039)
Female teacher (Reference: Male teacher)	-0.001
	(0.007)
Ln Years of experience	0.002
	(0.004)
Teachers do not have adequate instructional materials and supplies (Reference: Serious problem)	
Not a problem	0.038***
	(0.007)
Minor problem	0.044***
	(0.010)
Moderate problem	0.044***
	(0.009)
Disruptive students (Reference: A lot)	
Not applicable	0.039***
	(0.015)
Not at all	0.043***
	(0.015)
Some	0.033**
	(0.014)
Ln Minutes per week teaching reading	0.018*
	(0.010)
Sigma U (Constant)	-6.394***
	(0.315)
Sigma V (Constant)	-7.224***
	(0.328)
Constant	5.324***
	(0.112)
Observations	172

Source: Authors' own calculations from PIRLS and TIMSS 2011 data.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The practice of questioning –as asking for explanations– by teachers helps students to understand better the lessons and then enhance their achievement. Related to these results, Muijs and Reynolds (2010) found that time spent teaching interactively instead of using seat work or group work contributes to increase teacher effectiveness, although using this method in exclusive without any other procedure could have a negative effect on students' achievement. This interactive approach also includes engaging with students and asking them open questions, allowing multiple answers, etc (Gagne *et al.*, 1993). The value of these behaviors can be found in the need of students in primary education for explicit cognitive structuring (Stillings *et al.*, 1995). Bonesrønning (2003) deepened in other teacher behaviors and methodologies which would influence in their effectiveness. Among these variables there were: monitoring students' activities –positive influence–, summarizing the content and repeating explanations –positive and negative influence, respectively–.

From our results it can be seen that sending homework in half or more of the week's class days also aid students to reinforce the knowledge teachers are providing to them. A high homework assignment frequency was found to have a positive influence of students' achievement, as stated in works as Trautwein (2007), although not so homework time, because too much of it would show students' wrong routines, like few motivation or concentration (Trautwein and Köller, 2003). However, always discussing the homework does not seem to benefit students. This "discussing" involves a further correction of the homework, more than a mere process of correction and feedback. This effect could be due to the fact that the low experience and maturity of students to argument in a robust way might result in time lost, which could have been applied into advancing with the content of the subject.

Besides of the previous variables, teacher effectiveness is usually measured by using the standardized scores obtained by students in regional or national achievement tests. The fact that their salaries or other working advantages may be conditioned by this fact can make teachers who are not really effective –and thus, do not have proper methods to increase students' effort and learning– to be incorrectly classified as so (Rothstein *et al.*, 2010). Because of that, we have also controlled by the approach of "teaching to the test", which has been extended nowadays due to the increasing importance of international assessment tests as PISA, TIMSS or PIRLS in the so-called "teacher rankings" (Jürges and Schneider, 2008). Our results have shown that, when teachers are excessively focused on national tests, they may tend to orient their lessons to pass those tests more than making students learn.

However, as previously stated, the field of teacher efficiency has been treated by few studies from the point of view we are proposing in our research: Literature is mainly focused on efficient assignment of teachers' salaries and costs, more than on teachers' efficient use and interaction with classroom or school resources in order to carry out their lessons. Because of that, there is not a clear list of common used inputs, so we will define it by answering the question we have previously posed in section four. Table 2 presents the results for the estimates of teacher efficiency in reading.

As happened with effectiveness, a higher cultural level of the family increases students' achievement, although teachers' sex and years of experience do not affect students' achievement. The lack of an adequate provision of materials and having a lot of disruptive students may suppose a lower achievement for students taught by the teacher, as could be expected. Better school facilities and conditions are highlighted in the literature to be essential in order to make easier teachers' tasks, helping not only to improve their efficiency when performing their activities, but also their retention in the school (Leung *et al*., 2006). Muijs *et al*. (2014) indicated that teachers must manage and organize the classroom environment they have been provided as an efficient learning environment, in order to increase the engagement of their students (Creemers and Reezigt, 1996). It is also important that they manage efficiently interruptions and misbehaviors, so they have to deal with disruptive students –their assignment to the teacher class in not under his/her control– by establishing clear rules and procedures when the course begins, which must be precisely delimited and rigorously implemented (Creemers, 1994). Students' achievement also increases when minutes of teaching are higher, so more hours of class would foster a better control of the subject by students. Thus, the behavior of these variables shows that a favorable environment for teaching provides a better result of students.

Once obtained both estimations for effectiveness and efficiency, we proceed to sort the values of effectiveness and efficiency obtained in both specifications and then to obtain the 10% of top and low performers in each field. We have removed from the results those teachers who had a missing value for effectiveness or efficiency, due to the main objective of our paper: making a comparison between these characteristics. This will suppose to work with 156 teachers. Then, we denoted as "top performer" teachers those who ranked inside the 10% higher average reading score, efficiency or effectiveness –in each field there is a separate ranking–, and "low performers" for those in the 10% lower positions of the same rankings (i.e., we selected the 16 teachers in the top and 16 in the bottom in each of the three rankings).

One remarkable result is that only 44% of teachers with the higher average in reading scores also showed a top position in effectiveness and efficiency rankings[3]. However, for the lower positions, 75% of teachers with low average scores obtained the lowest positions in both effectiveness and efficiency rankings. Because of that, we can see that achievement is not the unique relevant variable to have into account when elaborating a ranking on teachers' capacities, so "league tables" will not be a proper approach, as Goldstein and Spiegelhalter (1996) and Leckie and Goldstein (2009) stated. About 6.25% of the teachers were allocated only in the low performers' rank in both effectiveness and efficiency (anyone in the top). In the case of effectiveness and average scores in reading, 12.5% of teachers were into the same top and low performers' ranking, and for the case of only average scores in reading and efficiency, a 12.5% (in the bottom of the rank).

6. Conclusions

As it was indicated, teacher effectiveness and efficiency are two desirable properties for education systems which have not been clearly defined in the previous literature and hence confused very frequently. To make clear the distinction we have proposed many instruments focused on analyzing the factors affecting both dimensions.

From our results on teachers' effectiveness it can be inferred that teachers who pose questions to primary school students when they are receiving the lessons are effective in order to make clearer the concepts which are being taught. In addition, frequent homework submissions have appeared to be a good procedure in order to make students reinforce the knowledge they are receiving in class, although a further discussing of this homework could be over the possibilities of the students at this age. Consequently, it is important to raise awareness among primary education teachers about the relevance of a frequent homework-control and questioning during the lessons, what can be achieved by the incorporation of these methods to the curriculum of teachers which are still in training process.

[3] Rankings were not included for reasons of space, but they will be provided upon request to the authors.

Teaching students by focusing on overcoming a particular external test has appeared to be a negative factor for students' learning. A potential solution for this is that teachers' salaries and compensations should not be determined by the ranking that they have obtained in these tests, because it could promote –even unconsciously– these practices.

In the case of teachers' efficiency, the availability of materials is essential for the proper development of the lessons, so a direct communication between directors and teachers has to be achieved, in order to make easier for teachers to ask the director about any needed material and for directors to make sure that teachers have the required materials for their lessons. Although having disruptive students is not under the control of the teacher, their curriculum should also be prepared to deal with these problems, as it was argued for effectiveness practices.

A relevant conclusion is that "league tables" are not a proper way of measuring the effectiveness or efficiency of a country, although they have been commonly used when international results are released (Leckie and Goldstein; 2009). Our results suggest that the development of such a ranking should be based on teacher effectiveness and efficiency characteristics, and not only on raw scores. These rankings can be used to determine their salaries or compensations in a fairer way, with the aim of motivating them to continue with their successful practices.

Although it could be argued that one of the limitations of our analysis is the use of cross-section data, many authors have declared that the results of cross-sectional analysis are as reliable as others obtained from panel data analysis (Lenkeit, 2012a) and could even overcome the latter (Teddlie *et al*., 2000).

Finally, a future line of research can be established in the use of the data from TIMSS 2015 to replicate the present study and to make comparisons in teacher level.

References

Arpino, B. and Varriale, R. (2009). *Assessing the quality of institutions' rankings obtained through multilevel linear regression models*. MPRA Paper No. 19873.

Beaman, R. and Wheldall, K. (2000). Teachers' Use of Approval and Disapproval in the Classroom. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *20(4)*, 431-446.

Bonesrønning, H. (2003). Can effective teacher behavior be identified? *Economics of Education Review*, *23*, 237-247.

Boonen, T., Van Damme, J. and Onghena, P. (2013). Teacher effects on student achievement in first grade: which aspects matter most? *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 25(1)*, 126-152.

Borich, G. (1996). *Effective teaching methods* (3rd ed.). New York: Macmillan.

Braga, M., Paccagnella, M. and Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review, 41*, 71-88.

Brophy, J. E. and Good, T. L. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328-375). New York: MacMillan.

Creemers, B. P. M. (1994). *The effective classroom.* London: Cassell.

Creemers, B. P. M. and Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*, 197-228.

De Leeuw, J. and Meijer, E. (2008). Introduction to multilevel analysis. In J. De Leeuw and E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 1-75). New York: Springer.

Dolton, P. and Marcenaro, O. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic Policy, 26(65)*, 5-55.

Foy, P. and O'Dwyer, L. M. (2013). *School Effectiveness Models and Analyses*. In M. O. Martin and I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011 Relationships report* (Technical Appendix B). United States: Boston College.

Gagne, R. M., Yekovich, C. W. and Yekovich, F. R. (1993). *The cognitive psychology of school learning*. New York, NY: HarperCollins.

Goldstein H. (2003). Multilevel modelling of educational data. In D. Courgeau (Ed.), *Methodology and epistomology of multilevel analysis* (pp.25-42). London: Kluwer.

Goldstein, H. and Healy, M. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society: Series A, 158*,175-177.

Goldstein H. and Spiegelhalter D. (1996). League tables and their limitations: statistical issues in comparisons of institutional performances. *Journal of the Royal Statistical Society: Series A, 159*, 385-443.

Grilli, L., Pennoni, F. Rampichini, C. and Romero, I. (2014). Exploiting TIMSS and PIRLS combined data: multivariate multilevel modelling of student achievement. *Presented at the VI European Congress of Methodology, 23-25 July 2014*. Utrecht, Netherlands.

Jacob, B. A. and Lefgren, L. (2008). Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education. *Journal of Labor Economics, 26(1)*, 101-136.

Jürges, H. and Schneider, K. (2008). Fair ranking of teachers. *Empirical Economics, 32*, 411-431.

Klassen, R. M. and Chiu, M. M. (2010). Effects on Teachers' Self-Efficacy and Job Satisfaction: Teacher Gender, Years of Experience, and Job Stress. *Journal of Educational Psychology, 102(3)*, 741-756.

Koopmans, T. (1951). Analysis of Production as an Efficient Combination of Activities. In T. C. Koopmans (Ed.), *Activity Analysis of Production and Allocation* (pp. 33-97). New Haven, Yale: University Press.

Lavy, V. (2002). Evaluating the effect of teachers group performance incentives on pupil Achievement. *Journal of Political Economy, 110*, 1286-317.

Lavy, V. (2004). *Performance pay and teachers' effort, productivity and grading ethics*. NBER Working Paper 10622.

Leckie, G. and Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A, 172(4)*, 835-851.

Lenkeit, J. (2012a). Effectiveness measures for crosssectional studies: a comparison of value-added models and contextualized attainment models. School Effectiveness and School Improvement: An International Journal of Research. *Policy and Practice, 24(1)*, 1-25.

Lenkeit, J. (2012b). How effective are educational systems? A value-added approach to measure trends in PIRLS. *Journal for Educational Research, 4*, 143-173.

Leung, M., Chan, J. K.W. and Wang, Z. (2006). Impact of school facilities on working behavior of teachers. *International Journal of Strategic Property Management, 10(2)*, 79-91.

Loeb, S. and M. Page (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics, 82(3)*, 393-408.

Manzi, J., Martín, E. S. and Bellegem, S. V. (2014). School system evaluation by value added analysis under endogeneity. *Psichometrika, 79(1)*, 130-153.

Marcenaro, O. (2002). *Los estudiantes universitarios y el uso de su tiempo: una función de producción educativa*. PhD Dissertation, Universidad de Málaga.

McEwan, P. J. and Carnoy, M. (2000). The Effectiveness and Efficiency of Private Schools in Chile's Voucher System. *Educational Evaluation and Policy Analysis, 22(3)*, 213-223.

Meeusen, W. and Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production function with composed errors. *International Economic Review, 18(2)*, 435-444.

Muijs, D., Kyriakides, L. and Greetje van der Werf, B. (2014). State of the art - teacher effectiveness and professional learning. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 25(2)*, 231-256.

Muijs, D. and Reynolds, D. (2010). School Effectiveness and Teacher Effectiveness in Mathematics: Some Preliminary Findings from the Evaluation of the Mathematics Enhancement Programme (Primary). *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 11(3)*, 273-303.

Nachtigall, C., Kröhne, U., Enders, U. and Steyer, R. (2008). Causal effects and fair comparison: Considering the influence of context variables on student competencies. In J. Hartig, E. Klieme and D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 315-335). Göttingen: Hogrefe and Huber.

Naper, L. R. (2010). Teacher hiring practices and educational efficiency. *Economics of Education review, 29(4)*, 658-668.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Raudenbush, S. and Willms, J. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*, 307-335.

Reynolds, D., Sammons, P., Stoll, L., Barber, M. and Hillman, J. (1996). School effectiveness and school improvement in the United Kingdom. *School Effectiveness and School Improvement, 7*, 133-158.

Rothstein, R., Ladd, H. F., Ravitch, D., Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Linn, R. L., Shavelson, R. J. and Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.

Scheerens, J. and Creemers, B. P. M. (1996). School effectiveness in the Netherlands: The modest influence of a research programme. *School Effectiveness and School Improvement,7*, 181-195.

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

Snijders, T. A. B. and Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publishers.

Stillings, N. A., Weisler, S. W., Chase, C. H., Feinstein, M. H., Garfield, J. L. and Risland, E. L. (1995). *Cognitive psychology*. Cambridge, MA: MIT Press.

Teddlie, C., Stringfield, S. and Reynolds, D. (2000). Context issues within school effectiveness research. In C. Teddlie and D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 160-186). London, UK: Routledge.

Thompson-Schill, S. L., Ramscar, M. and Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current directions in Psychological Science, 18(5),* 259-263.

Trautwein, U. (2007). The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction, 17(3),* 372-388.

Trautwein, U. and Köller, O. (2003). The relationship between homework and achievement: still much of a mystery**.** *Educational Psychology Review, 15*, 115-145.

Zuzovsky, R. (2009). Teachers' qualifications and their impact on student achievement: Findings from TIMSS 2003 data for Israel. In M. von Davier and D. Hastedt (Eds.), *IERI Monograph Series Issues and Methodologies, Vol. 2* (pp. 37-62). Hamburg/Princeton, Germany/USA: IERI Institute.