Identifying effective teachers: the case study of Spain

LUIS ALEJANDRO LOPEZ–AGUDO
Universidad de Málaga.
Facultad de Ciencias Económicas y Empresariales.
lopezagudo@uma.es

OSCAR MARCENARO–GUTIERREZ
Universidad de Málaga.
Facultad de Ciencias Económicas y Empresariales.
odmarcenaro@uma.es

The formation of effective teachers has been one of the main aims of education systems from their very beginning. The main objective of this research is to identify effective teachers based on their teaching practices and efficient use of their available resources. It is common in the effectiveness literature to rely on the achievement of teacher's students to measure the effectiveness of a particular teacher. Nevertheless, we propose the capacity of teachers to engage students in their lessons as an alternative to students' achievement, as it may be reflecting their ability to raise students' learning interest and, thus, denoting their effectiveness.

In order to indentify these teachers, we focus on fourth grade reading and mathematics teachers in Spain –using TIMSS and PIRLS 2011 data– and we propose a two step procedure whose first step consists of obtaining a measure of teachers' efficiency using their available educational resources to explain their students' achievement. Then, the latter is employed in a second step together with teachers' practices in the classroom to explain students' engagement in lessons.

The relevance of this research is rooted on the determination of the characteristics, teaching procedures and efficient use of resources which can make a teacher effective. To the extent that these effective teachers are able to improve their students' engagement thanks to their effective methods and independently of their students' socio-economic background, they would be fostering the improvement of social mobility and education systems, and hence the socio-cultural development of the Spanish society.

Keywords: teacher's effectiveness; teacher's efficiency; stochastic frontier analysis; multilevel analysis.

Acknowledgements: This work has been partly supported by the Andalusian Regional Ministry of Innovation, Science and Enterprise (PAI group SEJ-532 and Excellence research group SEJ-2727); the Research Plan of the University of Malaga (Capacity Building Programme I+D+i of Universities 2014-2015); by the Ministry of Economy and Competitiveness of Spain (Research Project ECO2014-56397-P) and the scholarship FPU2014 04518 of the Spanish Ministry of Education. Luis Alejandro Lopez-Agudo also acknowledges the training provided by the *Programa de Doctorado en Economía y Empresa de la Universidad de Malaga*.

1. INTRODUCTION

Effectiveness and efficiency are concepts commonly related to the world of business and productive processes. However, despite their relevance, they are usually confused in the economics of education literature (Lockheed & Hanushek, 1994). In this paper we focus on both concepts to shed light on the empirical delimitation of the factors and teaching methodologies –in the case of the Spanish education system– which could condition teachers' effectiveness, in which their efficient use of classroom resources might occupy a relevant place.

In the effectiveness literature it is common to rely on the achievement of students to measure the effectiveness of a particular teacher. Nevertheless, we use the capacity of teachers to engage students in their lessons as an alternative to student achievement, focusing on the definition of effective teachers proposed by Creemers and Reezigt (1996) and Kyriakides (2008) "effective teachers are expected (…) to maximize engagement rates". The idea is that so-defined effective teachers are expected to increase their students' interest and implication in their studies, thanks to their teaching practices. This way, we use students' engagement as the outcome of an effective teacher in order to provide useful information to orientate educational policy decisions.

With regard to the concept of efficiency included in effectiveness, it basically means getting the maximum output from the amount of inputs teachers use in the teaching-learning process, or achieving the same output with a minimum quantity of inputs[92]. The definition of Creemers and Reezigt (1996) and Kyriakides (2008) also highlights that effective teachers are expected "(…) to organize and manage the classroom environment as an efficient learning environment". Hence, teachers' efficiency might be defined as their ability to perform fruitful work in an organized way, using available resources and time without wasting any of them.

The development of a measure to identify effective teachers is a relevant issue, as the traditional "league tables", which are based only on the average score of teachers' students, do not take into account sampling variability and other sources of error (Leckie & Goldstein, 2009). A proper classification of teachers in terms of effectiveness is particularly relevant in order to reward effective teachers based on their performance fostering higher levels of student engagement and learning.

Regarding the methodology used for this analysis, a two-step procedure has been followed, which –to the best of our knowledge– also entails a novelty in the study of teacher effectiveness. Firstly, teachers' efficiency has been estimated by using stochastic frontier analysis (SFA), using as output students' academic achievement in reading and mathematics and, as inputs, available school resources for teachers. Then, the efficiency scores obtained from this procedure have been employed in a second step, together with teachers' learning practices in the classroom, to explain students' engagement in reading and mathematics

[92] Own adaptation of Lockheed and Hanushek (1994) –and references therein– general definition of efficient system.

lessons, by making use of multilevel regression models. To carry out this research, we made use of PIRLS (Progress in International Reading Literacy Study) and TIMSS (Trends in Mathematics and Science Study) combined databases in their 2011 wave for fourth grade –9/10 years old– Spanish students.

To sum up, this analysis of teachers' effectiveness allows us to determine whether efficiency is a relevant characteristic of effective teachers or not, together with other teaching practices. It will also help to establish teacher rankings that are more accurate and adapted to the pursued educational policy objectives than those based only on raw scores, enabling their use as educational policy instruments to define more precise educational interventions.

2. LITERATURE REVIEW

In general, the concepts of effectiveness and efficiency in teaching have been approached separately by previous literature (Lockheed & Hanushek, 1994). Beginning with the review of the literature related to effective teachers, teacher effectiveness has been frequently measured in the literature by using cross-sectional data –contextualized attainment models– from student and family characteristics (Nachtigall, Kröhne, Enders, & Steyer, 2008). However, when data for two years or more are available, most research proposes the formulation of value-added models (Manzi, Martín, & Bellegem, 2014) to measure the contribution that teachers make to students' prior achievement. Nevertheless, Lenkeit (2013), who focused on the subject of the differences between value-added models and contextualized attainment models, concluded that the latter are a proper alternative to value-added models, due to students' background characteristics acting as good substitutes for prior attainment. Moreover, this obstacle can also be overcome in our research by taking into account that Spanish students have the same teacher during a cycle[93], so their current learning and interest could be attributed to him/her (Jürges & Schneider, 2007). Additionally, the dependent variable which was chosen in our research to measure teacher effectiveness provides a proper solution to the problem of the lack of prior achievement: students report their degree of engagement with their current teacher –who is also the previous year's teacher, as students are in fourth grade–, so this variable would be directly related to the teacher's ability to generate his/her students' interest in the subject.

This ability of getting students engaged in the lesson has been highlighted in the literature as a distinctive characteristic of effective teachers. Timothy, Rupley, and Nichols (2007) stated that "Effective teachers not only use appropriate materials but also attend to actively engaging students in learning from the materials". In addition, Creemers and Reezigt (1996) denoted that effective teachers "are expected to organize and manage the classroom environment as an efficient learning environment and thereby to maximize engagement rates". These two definitions contain the main characteristics of an effective teacher we are considering in our

[93] Pursuant to law (art. 73 of the *Reglamento Orgánico de las escuelas de educación infantil y de los colegios de educación primaria*, Order 29th June, 1994), in 2011 –the date of the data under analysis– Spanish students had the same teacher for two years for each of the three cycles of primary education; in other words, they had three different teachers during their primary studies. Since we are studying fourth grade students, they should have been taught by the same teacher in third grade, so their engagement may be related to his/her teacher's performance.

research: a teacher who can use his/her learning materials and resources in an efficient way and, together with effective teaching methodologies, is able to engage students in the lesson and make them learn. The relevance of engaging students as a characteristic of an effective teacher has been remarked by some other authors, such as Taylor, Pearson, Clark, and Walpole (1999) for kindergarten to third grade United State students. The latter found that effective teachers engaged their students 96% of the time, while less effective teachers engaged their students only 63% of the time.

Focusing on teachers' efficiency, Ryan and Cooper (2007) claimed that the most efficient teachers could engage their students about thirty minutes per day more than average efficient teachers, and an hour more when compared to inefficient teachers. In this sense, Jürges and Schneider (2007) developed a stochastic frontier model in order to define a fair ranking of teachers based on their efficiency, employing German PIRLS data from 2001. Their results showed that less than two thirds of the ranked teachers continued in the same quality groups after controlling by student and school background characteristics, reaching the conclusion that rankings based only on raw test scores were unfair.

To the best of our knowledge, there is a virtual absence of literature which studies teacher effectiveness by also considering teacher efficiency as one of its elements at the Spanish primary education level, which is the gap we want to analyze.

3. DATA

Our analysis focuses on the effectiveness of the Spanish teachers who participated in TIMSS 2011 and PIRLS 2011. The purpose of TIMSS is to measure learning achievement in the areas of mathematics and science of students at the end of fourth (9-10 years) and eighth grades (13-14 years), while PIRLS is focused on reading achievement of fourth grade students. Fourth grade students were chosen in our study as they are more malleable at this age (Thompson-Schill, Ramscar, & Chrysikou, 2009) and would thus show a better reflection of teachers' procedures. PIRLS and TIMSS coincided in 2011 for the first time, so a total of 4,183 students who participated in both tests, and were taught by a total of 200 teachers, were chosen for our study. This sample was reduced in the efficiency analysis to 3,438 students taught by 184 teachers –for the case of reading– and 3,432 students taught by 184 teachers –for mathematics–, due to the information missing for some variables related to efficiency. Later, in the effectiveness analysis, after controlling for the predicted efficiency scores, we end up with a sample of 3,191 students for reading –taught by 156 teachers– and 3,201 students for mathematics –taught by 158 teachers–.

Students' engagement –proxy of teachers' effectiveness– is represented by an index provided by PIRLS 2011 –for reading– and TIMSS 2011–for mathematics–, called "Students Engaged in Reading/Mathematics Lessons index"[94]. We have classified each of the available variables

[94] Following Mullis, Martin, Foy, and Arora (2012), this index was created by asking students the degree to which they agreed ("Agree a lot"; "Agree a little"; "Disagree a little"; "Disagree a lot") with five statements: "I know what my teacher expects me to do", "I think of things not related to the lesson", "My teacher is easy to understand", "I am interested in what my teacher says" and "My teacher gives me interesting things to do"[94].

according to their adscription to teacher effectiveness or efficiency characteristics –supporting this classification in the previous literature–. When determining whether a teacher variable represents his/her effectiveness or efficiency, we have asked ourselves the following two questions: "Does this variable represent the available resources for teachers in order to develop their lessons, as materials, classroom characteristics, available time or students' background, which are not easily alterable by the teacher –efficiency–?"; "Does this variable describe a method or procedure defined by the teacher when teaching his/her students to make them exert effort in learning –effectiveness–?" The selected variables which respond to one of these questions and had a significant effect on explaining teacher efficiency or effectiveness were included in the estimations of Table 1 –efficiency– or Table 2 –effectiveness–. The main descriptive statistics of the selected variables are presented in Table A.1 –efficiency– and Table A.2 –effectiveness–, both displayed in Appendix[95].

4. METHODOLOGY

As indicated in the introduction section, the first step of this analysis –efficiency– begins with the prediction of teachers' efficiency by using stochastic frontier analysis, considering as output students' achievement and as inputs teachers' available learning resources. This efficiency measure is employed in the second step of the analysis –effectiveness– as a regressor, together with some teacher characteristics, learning practices and procedures, which will explain students' engagement.

4.1 First step: Stochastic frontier analysis

In the first step of our model, we make use of stochastic frontier analysis to obtain a measure of teacher efficiency. Concretely, a time-invariant stochastic frontier model (Battese & Coelli, 1988) has been proposed. It helps us define a parametric function in which there are two main levels: the first level is students, whose units –students– are grouped in the second level –teachers–. This approach is more comparable with the two-level parametric form of the multilevel regression model than other non-parametric approaches. In addition, measurement errors are a major concern when dealing with students' score data, which is a field where SFA outperforms non-parametric approaches. This procedure has the particularity that it presents a composed error term, which is the sum of a measurement or specification error and a one-sided disturbance, which represents inefficiency. We have employed a log-linear Cobb-Douglas education function, in line with Chakraborty, Biswas, and Lewis (2001), who remark that its use would be more appropriate than other forms as translog when there are fewer degrees of freedom –as in our case–. This frontier model can show the best combination of inputs that maximize average students' performance. Formally:

[95] Many other variables related to teachers' methodologies were analyzed –as they fulfilled the criteria of the second question–, but they did not result significant to explain teacher effectiveness. The complete list will be provided upon request to the authors.

$$\ln(y_{ji}) = \alpha + \sum_{k=1}^{p} \beta_k \ln(x_{kji}^{(1)}) + \sum_{s=1}^{l} \gamma_s \ln(z_{sj}^{(1)}) + v_{ji} - u_j$$

$$v_{ji} \sim N(0, \sigma_v^2)$$

$$u_j \sim F$$

For $j = 1, \dots, n$ teachers and $i = 1, \dots, m$ students. (1)

where y_{ji} denotes the output (scores in reading or mathematics) of student "i" taught by teacher "j"; $x_{kji}^{(1)}$ is the k^{th} input of the socio-economic background characteristics of student "i" –taught by teacher "j"– and β_k its technology parameter, while $z_{sj}^{(1)}$ represents the s^{th} input of teacher learning resources –time, materials, etc.– of teacher "j" and γ_s its technology parameter. The composite error consists of the sum or the difference of the disturbance u_j (representing inefficiency) –as it is a time-invariant model $u_{ji} = u_j$– and v_{ji}, which is a normally distributed disturbance, these two terms being independent of each other and $i.i.d.$ between observations. After performing our analysis we obtained a better fit for distribution (F) of the inefficiency term u_j with the truncated-normal one –$u_j \sim N^+(\mu, \sigma_u^2)$– (also following Battese & Coelli, 1988). Due to the assumptions to identify the inefficiency term, the estimation has to be performed by maximum likelihood. Once this model has been estimated, we predict a measure of efficiency –u_j– for each teacher.

4.2 Second step: Multilevel regression analysis

Standard regression models do not take into account the hierarchical structure which educational data usually presents. Because of that, and following works such as Lenkeit (2013), multilevel regression analysis (Hierarchical Linear Models, HLM) constitutes a proper approach in order to study effectiveness when cross-section data are available and is thus used in the second step of our analysis. This methodology has been widely used due to its capacity to account for the existence of a hierarchical organization of data in the teaching-learning process (Snijders & Bosker, 2012). Specifically, we 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 is defined for each teacher in the second level. The estimated model can be specified in a general way:

$$E_{ij} = \lambda_0 + \sum_{h=1}^{q} \delta_h x_{hij}^{(2)} + \sum_{r=1}^{g} \vartheta_r z_{rj}^{(2)} + w_j + \varepsilon_{ij}$$

For $i = 1, \dots, m$ students and $j = 1, \dots, n$ teachers. (2)

where E_{ij} denotes the engagement index in the subjects of reading or mathematics of student "i" who is taught by teacher "j"; $x_{hij}^{(2)}$ reflects the h^{th} characteristic –for $h = 1, \dots, q$– of student "i" –taught by teacher "j"–; δ_h is the slope for the relationship between E_{ij} and $x_{hij}^{(2)}$; $z_{rj}^{(2)}$ reflects the r^{th} characteristic or methodology –for $r = 1, \dots, g$– of teacher "j"; ϑ_r is the

slope for the relationship between E_{ij} and $z_{rj}^{(2)}$; ε_{ij} is the idiosyncratic error term, which is assumed independent among students and also independent of w_j. It is assumed that each teacher has a different level parameter (interception term), which is not fixed and behaves as a random variable with an average value (λ_0) affected by deviations represented by the random variable w_j.

In order to establish a quantitative measure of teachers' effectiveness, teacher-level residuals have been used, because they can be defined as predictions of the random effects representing teacher effectiveness. Following Raudenbush and Willms (1995), these residuals can be obtained by using the Empirical Bayes (EB) method, which produces the so-called "shrunken residuals" –also known as best linear unbiased predictors (BLUPs)–. They are shrunk but efficient estimates of teacher effectiveness and avoid the accidental assignment of a teacher from the bottom to the top of the ranking (Arpino & Varriale, 2010), although they are affected by the sampling variability and other errors. Because of this, Goldstein and Healy (1995) suggested creating pairwise confidence intervals in order to overcome this uncertainty.

5. RESULTS

The main results obtained from both teacher efficiency and effectiveness analysis for the subjects of reading and mathematics, performed for the selected variables in each step, are discussed below.

5.1 First step: Efficiency analysis

Previous literature focuses mainly on the 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, the field which our study deals with. Table 1 presents the results for the estimates of teacher efficiency in reading and mathematics. These efficiency estimations have been made by considering students' achievement in reading and mathematics as dependent variables.

Table 1. Teacher efficiency estimates for reading and mathematics

Variables	Ln(Reading Scores)	Ln(Mathematics Scores)
Student level variables		
Ln (Home Resources for Learning index)	0.223***	0.282***
	(0.012)	(0.013)
Student with diglossia (Reference: Student without diglossia)	-0.040***	-0.033***
	(0.007)	(0.007)
Teacher level variables		
Classrooms are overcrowded (Reference: Serious problem)		
Not a problem	0.002	0.040***
	(0.012)	(0.015)

Variables	Ln(Reading Scores)	Ln(Mathematics Scores)
Minor problem	-0.005	0.034*
	(0.012)	(0.018)
Moderate problem	0.007	0.038**
	(0.013)	(0.015)
Teachers do not have adequate instructional materials and supplies (Reference: Serious problem)		
Not a problem	0.062**	0.058
	(0.024)	(0.041)
Minor problem	0.067***	0.048
	(0.024)	(0.042)
Moderate problem	0.061**	0.061
	(0.024)	(0.041)
Minutes per week teaching reading/mathematics (Reference: 240 or less)		
More than 240 to 300	0.023***	0.009
	(0.009)	(0.009)
More than 300	0.013	0.012
	(0.008)	(0.013)
Constant	5.706***	5.526***
	(0.039)	(0.058)
Observations: Students	3 438	3 432
Groups: Teachers	184	184
Wald test	416.61***	517.19***

Source: Authors' own calculations from PIRLS and TIMSS 2011 data.

Dependent variable: Ln(Students achievement in reading/mathematics in PIRLS/TIMSS). Range of the variables: Ln(Reading Scores): 5.72 – 6.53 (Original variable range: 304 – 687); Ln(Mathematics Scores): 5.50 – 6.52 (Original variable range: 245 – 677).

Estimation method: Time-invariant stochastic frontier. The efficiency term is supposed to follow a truncated normal distribution.

Standard errors in parentheses.

*** denotes variable significant to level 1%; ** to 5%; * to 10%.

Focusing on human resources –i.e. students– teachers work with, we make use of the Home Resources for Learning index provided by PIRLS 2011[96], as this variable is a good "proxy" of students' background characteristics. It was found that those teachers who have students from high socio-economic backgrounds are able to obtain better results from them in reading and mathematics (Calero, Choy, and Waisgrais, 2010; Marcenaro, 2014). The fact that students present diglossia is detrimental for students' achievement in both subjects –students who

[96] This index is derived from three home questionnaire variables –parents' education, parents' occupation and number of children books at home– and two children questionnaire variables –number of books in the home and availability of key study supports at home: an internet connection and their own room–.

present it perform 4% and 3.3% lower in reading and mathematics, respectively, than those who do not–. A similar relationship was found by Carabaña (2013).

In the case of material resources, when the lack of an adequate provision of materials is not a serious problem, it may entail higher achievement in reading for students taught by the teacher –approximately 6%–. Better school facilities and conditions are highlighted in the literature as essential in order to make teachers' tasks easier, helping them not only improve their efficiency when performing their activities, but also their retention in the school (Leung, Chan, & Wang, 2006). However, this lack of materials does not seem to affect scores in mathematics. Our results show that when overcrowded classrooms are not a serious problem they may have a positive effect –4% on average– on scores in mathematics. Blatchford and Martin (1998) found that primary school students in the United Kingdom received better support in smaller classes than in overcrowded ones, and that teacher-pupil interactions were higher in smaller ones. Finally, for time resources, students' achievement in reading is increased 2.3% when teachers teach between 4 and 5 hours in a week –compared to 4 hours–, although a higher number of hours does not seem to have a significant effect.

To sum up, the behavior of these variables shows that a favorable environment for teaching may help teachers obtain better results from their students.

5.2 Second step: Effectiveness analysis

Once an efficiency score prediction for reading and mathematics teachers is obtained from the efficiency estimates in Table 1, this efficiency measure is included as a regressor in the estimations presented in Table 2, where teachers' effectiveness, proxied by students' engagement in reading and mathematics lessons, is analyzed.

Table 2. Teacher effectiveness estimates for reading and mathematics

Variables	Engagement in Reading	Engagement in Mathematics
Student level variables		
Female student (Reference: Male student)	0.431***	0.038
	(0.066)	(0.068)
Teacher level variables		
Female teacher (Reference: Male teacher)	0.281*	0.097
	(0.149)	(0.138)
Years of experience	0.011**	0.011**
	(0.005)	(0.005)
Efficiency	0.809***	0.766***
	(0.290)	(0.275)
Discussions with other teachers about how to teach a concept (Reference: Never)		
Daily or almost daily	0.502**	1.002***
	(0.246)	(0.214)

Variables	Engagement in Reading	Engagement in Mathematics
1 to 3 times a week	0.106	0.460**
	(0.221)	(0.184)
2 or 3 times a month	0.131	0.546***
	(0.218)	(0.180)
Visit another teacher's classroom (Reference: Never)		
Daily or almost daily	-0.589	-1.571***
	(0.572)	(0.454)
1 to 3 times a week	-0.291	-0.216
	(0.203)	(0.182)
2 or 3 times a month	0.020	0.018
	(0.159)	(0.150)
Monitor whether or not the homework was completed: Always or almost always (Reference: Sometimes, Never or almost never)	0.659**	-0.333
	(0.282)	(0.258)
Discuss the homework in class (Reference: In reading: Never or almost never; In mathematics: Sometimes, Never or almost never)		
Always or almost always	1.095**	0.064
	(0.436)	(0.135)
Sometimes	1.304***	-
	(0.435)	-
National achievement tests (Reference: Little or no emphasis)		
Major emphasis	0.275	-0.134
	(0.233)	(0.195)
Some emphasis	0.118	0.286**
	(0.122)	(0.115)
Take a written test or quiz (Reference: In reading: Never; In mathematics: Some lessons, Never)		
Every or almost every lesson	0.147	0.388*
	(0.418)	(0.222)
About half the lessons	-0.161	0.448**
	(0.255)	(0.186)
Some lessons	0.042	-
	(0.243)	-
Work on problems (individually or with peers) while I am occupied with other tasks (Reference: Never)		
Every or almost every lesson	-	-0.008
	-	(0.164)
About half the lessons	-	-0.099
	-	(0.171)
Some lessons	-	-0.281**
	-	(0.137)
Read aloud to the class (Reference: Once or twice a month, Never or almost never)		
Every or almost every day	-0.675***	-

Variables	Engagement in Reading	Engagement in Mathematics
	(0.252)	-
Once or twice a week	-0.647**	-
	(0.282)	-
Constant	7.289***	9.027***
	(0.570)	(0.371)
Observations: Students	3 191	3 201
Groups: Teachers	156	158
Wald test	105.08***	78.66***
Snijders/Bosker R-squared student level	0.061	0.053
Snijders/Bosker R-squared teacher level	0.274	0.313
ICC	0.086***	0.063***
LR test for linear regression	121.28***	71.65***

Source: Authors' own calculations from PIRLS and TIMSS 2011 data.
Dependent variable: reading/mathematics lessons engagement index –index data from students– in PIRLS/TIMSS. Range of the variables: Engagement in Reading: 2.07 – 14.32; Engagement in Mathematics: 2.68 – 13.27.
Estimation method: Multilevel regression.
Standard errors in parentheses.
*** denotes variable significant to level 1%; ** to 5%; * to 10%.

Firstly, teachers' general characteristics and those not related to their teaching methodologies are analyzed[97]. In the case of reading, the fact that students are girls may suppose an increase in their engagement compared to boys, which could also be reflected in that female teachers are able to make their students more engaged in their reading lessons than male teachers, maybe due to the transmission of their enjoyment of this subject to their students. Furthermore, teachers' years of experience have appeared to be positive to engage students in lessons. A non-linear relationship of students' engagement index and years of experience was checked, finding that it is not accomplished. In the case of efficiency, an increase in teachers' efficiency score[98] of one normalized point implies a higher index of teacher engagement of approximately 0.8 in both subjects. These variables are often accompanied by others which account for teacher's highest level of education. Nevertheless, in the case of our data, 153 out of 156 reading teachers (154 of 158 for mathematics teachers) had ISCED level 5a (first), and only 3 (4 for mathematics) had ISCED level 5a (second), making it irrelevant to control for this characteristic.

Dealing with the effectiveness of teachers' methodologies, the discussion with other teachers on how to teach a concept in reading –only for daily or almost daily frequencies– and mathematics is positive for students' engagement. The higher effect in mathematics may be explained by the higher complexity of this subject, which could require further discussion and

[97] There are variables which present a low number of observations in some categories. Because of that, these categories have been grouped with others. The description of this grouping has been included in the footnotes of Table A.2 (Appendix).

[98] These efficiency scores have been re-scaled following the formula for a "j" efficiency score –for teacher "j"– in subject c ($c = 1$ for reading and $c = 2$ for mathematics): $EfficiencyScore_{cj} - MinimumEfficiencyScore_c)/(MaximumEfficiencyScore_c - MinimumEfficiencyScore_c)$. This procedure was employed in order to make the coefficients easily interpretable, as the range of the efficiency scores in reading is 83.62 – 99.24, while for mathematics it is 82.21 – 98.85.

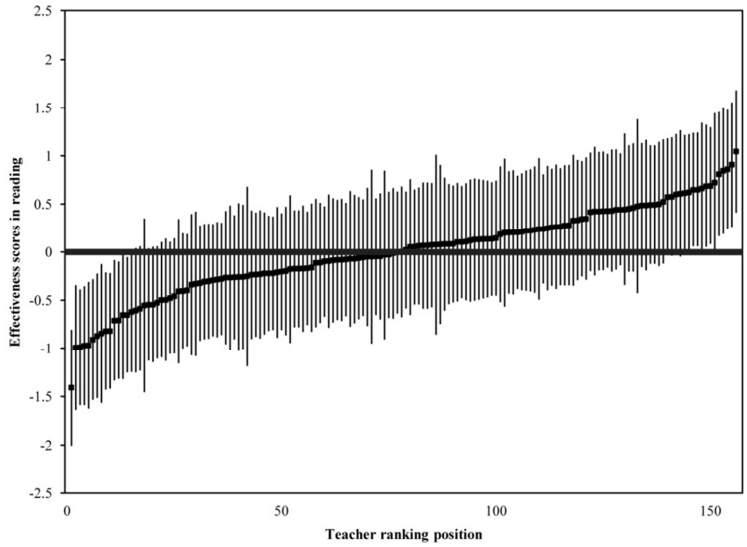
greater teacher consensus on how to approach a certain topic. The positive effects of teachers' collaboration in the definition of concepts and students curriculum has been highlighted by authors such as Grossman, Wineburg, and Woolworth (2001), who stated that this kind of teacher collaboration foments the intellectual renewal of the school, which in turn translates into learning benefits for students. In the case of daily or almost daily visits to other teachers' classrooms in mathematics, it has a negative effect on students' engagement, which could be due to the higher propensity of students in these classes to become distracted.

The frequent discussion in class of students' homework has a positive effect on engagement in reading, as reading activities may have different interpretations depending on the reader's point of view, while mathematics, as a deterministic subject, does not need a deep discussion of its results. Related to this issue, Muijs and Reynolds (2000) found that time spent teaching interactively instead of using seat work or group work contributes to increasing teacher effectiveness. Furthermore, the control of homework by teachers in reading is also positive for student engagement –increasing engagement by 0.659–, as indicated by authors as Bonesrønning (2004).

We have also controlled for the approach of "teaching to the test", which has been extended currently due to the increasing importance of international assessment tests in the so-called "teacher rankings" (Jürges & Schneider, 2007). However, our results have shown that some emphasis of teachers in national achievement tests increases their students' engagement in mathematics –by 0.286 points–, as well as the frequent control of students with a written test or quiz –by 0.388 for every or almost every lesson and 0.448 for about half the lessons–, which could increase the interest of students in paying attention in class, as mathematics may require a greater effort by students during lessons in order to understand certain concepts.

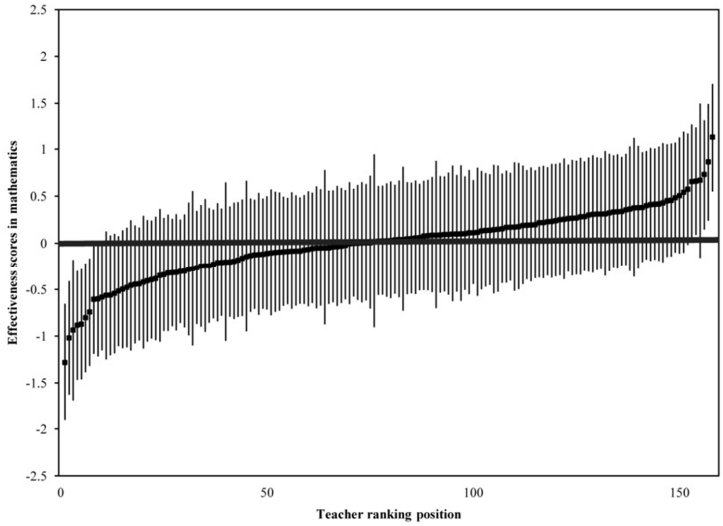
In the case of the methodologies which are more related to a particular subject –which is why we only have information about them for one subject– it appears that when students are not used to working on problems while their teacher is doing other tasks –which only reduces engagement in mathematics by 0.281 when done in some lessons–, they may lose interest in the subject. This has been highlighted by Muijs and Reynolds (2000), who stated that teachers should actively monitor students' individual and group work, being approachable and going around the classroom, instead of sitting at his/her desk. In the reading subject the practice of making students read aloud to the class has shown a negative effect on engagement, which could be due to a lack of interest of the other students who are not following the reading. According to Tindal, Heath, Hollenbeck, Almond, and Harniss (1998), reading aloud does not permit self-pacing for students, which undermines students' attention and, then, engagement.

Figure 1. Caterpillar plots of teacher effectiveness scores in reading



Note: "Bottom teacher" positions are located on the left side of the figure, while "top teacher" positions are on the right side.
Source: Authors' own calculations from PIRLS and TIMSS 2011 data.

Figure 2. Caterpillar plots of teacher effectiveness scores in mathematics



Note: "Bottom teacher" positions are located on the left side of the figure, while "top teacher" positions are on the right side.
Source: Authors' own calculations from PIRLS and TIMSS 2011 data.

Figures 1 and 2 show caterpillar plots of teachers' effectiveness scores in reading and mathematics –sorted in an ascending trend from the left, representing a "ranking" of effectiveness–, whose confidence intervals have been calculated at a level of 95%. As can be seen, most confidence intervals of effectiveness scores in the middle of the distribution

overlap with the horizontal line of the average of BLUPs[99] –showing the existence of little difference between teachers who obtained an "average" score–. There are 14 (9%) reading teachers and 10 (6%) mathematics teachers who perform significantly lower than the average, so we denote them as "bottom teachers". However, there are 13 (8%) reading teachers and 6 (4%) mathematics teachers who perform over the average, so they are called "top teachers". In addition, the predicted engagement that teachers achieve shows a clear higher mean for top teachers –10.54 for reading and 10.35 for mathematics– than "bottom teachers" –9.06 for reading teachers and 8.95 for mathematics teachers–. Hence, the structure of these figures may indicate that effective and non effective teachers present differences in the way they engage their students in lessons and, thus, the work of those teachers who are more effective should be recognized.

6. DISCUSSION AND CONCLUSIONS

The purpose of this research was to delimit, both theoretically and empirically, the teaching methodologies and efficient practices affecting teachers' effectiveness, focusing on Spanish fourth grade students. In order to achieve this aim, a novel combination –to the best of our knowledge– of techniques to measure teachers' efficiency and effectiveness was employed.

In the case of teachers' efficiency, a relevant conclusion is that the proper use of the available resources that this concept entails is essential to engage students in lessons and, thus, increase teachers' effectiveness. In this sense, the availability of sufficient teaching materials may be necessary for the proper development of lessons, so direct communication regarding this subject between head teachers and teachers should be achieved. Overcrowded classrooms seem to be a problem in mathematics lessons, so the number of students per class should be controlled and the alternative of splitting classrooms into groups may be advisable. Finally, an increase in the number of teaching hours may be positive for students, up to a certain threshold –5 hours a week–. The new education act in Spain –LOMCE (*Ley Orgánica de Mejora de la Calidad Educativa*)– has introduced new regulations on the latest practice.

From our results on teachers' effectiveness it can be inferred that interactions between teachers on how to explain a concept in a lesson may be positive, while visiting other teachers' classrooms is not. This might mean that teachers should define the contents of their subjects with the aid and collaboration of other teachers, but from the moment the class is in the hands of the teacher in charge of the classroom, he/she should manage his/her students without the intervention of another teacher. Results on homework may indicate that effective reading teachers usually discuss homework's content with the students in order to provide clarifications, and they also monitor its completion. These actions, together with putting some

[99] This "overlapping" can be easily observed if we focus on the horizontal line departing from the "0" value –the average of all BLUP scores–. Those teachers placed in the bottom positions of the ranking –left side of the figures– and whose upper bounds of their confidence intervals do not touch the average line can be considered as significantly different from "average teachers", so they are denoted as "bottom teachers". Similarly, those teachers in the high ranking positions –right side of the figures– whose lower bounds of their confidence intervals do not touch the average line can be considered as significantly different from "average teachers", and hence as "top teachers".

emphasis on national achievement tests –which may mean teaching students in a more mathematic "competence" than "curricular" manner– and making students take a written quiz in mathematics every or almost every lesson –which can help students to retain concepts more easily in a subject which is difficult for many of them– should be included in teachers' curriculum, as they have been found to be effective teaching practices. In addition, reading aloud to the class could provoke a decrease in students' attention to the reading, as they are not able to read at their own pace and understand the content and sense of the text. Furthermore, making students work on problems on their own without paying enough attention or not guiding them during the process could be detrimental for their engagement, as students may lose interest in the task.

Hence, as efficiency is integrated in the concept of effectiveness, putting more stress on making teachers manage their resources in a proper way would also help them to make their students become more interested and exert an effort in learning. The proposed classification of teachers in rankings according to their effectiveness can be used to develop educational policies in a fairer way than only taking into account "raw scores", and to determine teachers' salaries or compensations, with the aim of motivating teachers to continue with their successful practices. To the extent that these effective teachers are able to improve their students' engagement, thanks to their effective methods and independently of their students' socio-economic backgrounds, they are fostering the improvement of social mobility and educational systems.

REFERENCES

Arpino, B., & Varriale, R. (2010). Assessing the quality of institutions' rankings obtained through multilevel linear regression models. Journal of Applied Economic Sciences, 11(1), 7–22.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics, 38*(3), 387–399.

Blatchford, P., & Martin, C. (1998). The effects of class size on classroom processes: "It's a bit like a treadmill– working hard and getting nowhere fast". *British Journal of Educational Studies, 46*(2), 118–137.

Bonesrønning, H. (2004). Can effective teacher behavior be identified? *Economics of Education Review*, *23*(3), 237–247.

Calero, J., Choy, A., & Waisgrais, S. (2010). Determinantes del riesgo de fracaso escolar en España: una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006. *Revista de Educación, Extraordinary*, 225–256.

Carabaña, J. (2013). Repetición de curso y puntuaciones pisa ¿cuál causa cuál? In INEE (Ed.), *PISA 2012: Programa para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario* (pp. 32–66). Madrid: Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa.

Chakraborty, K., Biswas, B., & Lewis, W. C. (2001). Measurement of Technical Efficiency in Public Education: A Stochastic and Nonstochastic Production Function Approach. *Southern Economic Journal, 67*(4), 889–905.

Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 7*(3), 197–228.

Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society: Series A (Statistics and Society)*, *158*(1), 175–177.

Grossman, P., Wineburg, S., & Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College record, 103*(6), 942–1012.

Jürges, H., & Schneider, K. (2007). Fair ranking of teachers. *Empirical Economics, 32*(2), 411–431.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 19*(4), 429–446.

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

Lenkeit, J. (2013). Effectiveness measures for cross-sectional 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.

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

Lockheed, M., & Hanushek, E. (1994). Concepts of Educational Efficiency and Effectiveness. In R. Husén, & T. N. Postlethwaite (Eds.), *International Encyclopedia of Education* (pp. 1779–1784). Oxford: Pergamon.

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

Marcenaro, O. (2014). *El rendimiento del alumnado andaluz a examen*. Sevilla: Consejería de Presidencia e Igualdad de la Junta de Andalucía.

Muijs, D., & Reynolds, D. (2000). 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.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center, Boston College.

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

Order 29th June (1994). *Reglamento Orgánico de las escuelas de educación infantil y de los colegios de educación primaria*. Spain: BOE nº 160, 6th July 1994, pp. 21589–21597. https://www.boe.es/diario_boe/txt.php?id=BOE-A-1994-15723

Raudenbush, S. W., & Willms, J. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*(4), 307–335.

Ryan, K., & Cooper, J. (2007). *Those who can, teach*. New York: Houghton Mifflin.

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

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (1999). *Beating the odds in teaching all children to read: Lessons from effective schools and exemplary teachers*. Michigan: Center for the Improvement of Early Reading Achievement.

Thompson-Schill, S. L., Ramscar, M., & 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.

Timothy R. B., Rupley, W. H., & Nichols, W. D. (2007). The Effective Teacher of Reading: Considering the "What" and "How" of Instruction. *The Reading Teacher, 60*(5), 432–438.

Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children, 64*(4), 439–450.

APPENDIX

Table A.1. Descriptive statistics for the efficiency analysis

		Reading			Mathematics		
		Observations	Mean/Proportion	S.d.	Observations	Mean/Proportion	S.d.
	Student level variables						
Home Resources for Learning index		3,438	10.44	1.74	3,432	10.45	1.74
Student with diglossia	Yes	3,438	0.082	0.274	3,432	0.086	0.280
	No	3,438	0.918	0.274	3,432	0.914	0.280
	Teacher level variables						
Classrooms are overcrowded	Not a problem	184	0.364	0.482	184	0.370	0.484
	Minor problem	184	0.277	0.449	184	0.272	0.446
	Moderate problem	184	0.272	0.446	184	0.272	0.446
	Serious problem	184	0.087	0.283	184	0.087	0.283
Teachers do not have adequate instructional materials and supplies	Not a problem	184	0.451	0.499	184	0.451	0.499
	Minor problem	184	0.342	0.476	184	0.342	0.476
	Moderate problem	184	0.196	0.398	184	0.196	0.398
	Serious problem	184	0.011[a]	0.104	184	0.011[a]	0.104
Minutes per week teaching reading/mathematics	240 or less	184	0.190	0.394	184	0.375	0.485
	More than 240 to 300	184	0.337	0.474	184	0.495	0.501
	More than 300	184	0.473	0.501	184	0.130	0.338

Source: Authors' own calculations from PIRLS and TIMSS 2011 data.
[a] The number of students of these teachers is 47, so it is a proper reference category.

Table A.2. Descriptive statistics for the effectiveness analysis

		Reading			Mathematics		
		Observations	Mean/Proportion	S.d.	Observations	Mean/Proportion	S.d.
	Student level variables						
Sex of the student	**Male**	3,191	0.510	0.500	3,201	0.512	0.500
	Female	3,191	0.490	0.500	3,201	0.488	0.500
	Teacher level variables						
Sex of the teacher	**Male**	156	0.212	0.410	158	0.228	0.421
	Female	156	0.788	0.410	158	0.772	0.421
Years of experience	**14 years old or less**	156	0.346	0.477	158	0.329	0.471
	15 to 27 years old	156	0.321	0.468	158	0.335	0.474
	28 years old or more	156	0.333	0.473	158	0.335	0.474
Discussions with other teachers about how to teach a concept	**Daily or almost daily**	156	0.167	0.374	158	0.158	0.366
	1 to 3 times a week	156	0.353	0.479	158	0.361	0.482
	2 or 3 times a month	156	0.372	0.485	158	0.354	0.480
	Never	156	0.109	0.313	158	0.127	0.334
Visit another teacher's classroom	**Daily or almost daily**	156	0.013[a]	0.113	158	0.019[a]	0.137
	1 to 3 times a week	156	0.096	0.296	158	0.101	0.303
	2 or 3 times a month	156	0.167	0.374	158	0.165	0.372
	Never	156	0.724	0.448	158	0.715	0.453
Monitor whether or not the homework was completed	**Always or almost always**	156	0.955	0.208	158	0.949	0.220
	Sometimes	156	0.038	0.193	158	0.051	0.220
	Never or almost never	156	0.006[b]	0.080	158	0.000[b]	0.000
Discuss the homework in class	**Always or almost always**	156	0.744	0.438	158	0.766	0.425
	Sometimes	156	0.237	0.427	158	0.228	0.421
	Never or almost never	156	0.019[c]	0.138	158	0.006[d]	0.080
National achievement tests	**Major emphasis**	156	0.077	0.267	158	0.095	0.294
	Some emphasis	156	0.538	0.500	158	0.506	0.502
	Little or no emphasis	156	0.385	0.488	158	0.399	0.491
Take a written test or quiz	**Every or almost every lesson**	156	0.032	0.177	158	0.070	0.255
	About half the lessons	156	0.250	0.434	158	0.095	0.294
	Some lessons	156	0.635	0.483	158	0.829	0.378
	Never	156	0.083	0.277	158	0.006[e]	0.080
Work on problems (individually or with peers) while I am occupied with other tasks	**Every or almost every lesson**	-	-	-	158	0.196	0.398
	About half the lessons	-	-	-	158	0.171	0.378
	Some lessons	-	-	-	158	0.380	0.487
	Never	-	-	-	158	0.253	0.436
Read aloud to the class	**Every or almost every day**	156	0.788	0.410	-	-	-
	Once or twice a week	156	0.154	0.362	-	-	-
	Once or twice a month	156	0.051	0.221	-	-	-
	Never or almost never	156	0.006[f]	0.080	-	-	-

Source: Authors' own calculations from PIRLS and TIMSS 2011 data.

[a] The number of students of these teachers is 32 in reading and 41 in mathematics, so it is a proper standalone category.

[c] The number of students of these teachers is 75, so it is a proper reference category.

[b, d, e, f] This category has been grouped with the previous one in estimations in order to solve the problem of its small number of observations.