The version of record of this article has been published and is available in
Social Indicators Research (date of publication 25 October 2015)
https://doi.org/10.1007/s11205-015-1162-8

Balancing teachers' math satisfaction and other indicators of the education system's performance

Marcenaro-Gutierrez, O. D.[1,2], Luque, M.[3], and Lopez-Agudo, L. A.[2]

[2]*Department of Applied Economics (Statistics and Econometrics)*

[3] *Department of Applied Economics (Mathematics)*

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

Abstract

We investigate the potential balance between some teacher characteristics, particularly teachers' satisfaction and different measures of pupils' performance (average students' test scores and percentage of students achieving basic standards of learning in mathematics), in order to optimize the outputs of the Spanish education system. Our contribution to the existing literature is twofold: on the one side, we provide estimates on the balance between teachers' effectiveness and fourth grade students' performance in Spain, by using recent survey data from the program on Trends in International Mathematics and Science Study; on the other side, we implement a novel methodology which allows optimizing simultaneously a set of indicators of the educational system outputs, to the extent that our empirical approach revealed the existence of some degree of conflict among the outputs under scrutiny.

These analyses provide empirical evidence of the importance of simultaneously analysing different indicators of the performance of the education system and the need to invest in teachers' satisfaction as a motivation mechanism for improving national educational achievement, at least in primary education. Additionally they bring attention into the harmful impact of some education policy measures taken as a consequence of the economic crisis, as, e.g., the delay in the replacement of teachers, which entails the lack of teachers at schools to cover all lessons.

Key words: teachers' satisfaction, pupil performance, mathematics, trade-off, TIMSS 2011.

JEL code: I21, C61, J28.

1. Introduction

[1] Corresponding author e-mail: odmarcenaro@uma.es; Tel.: +34 95 213 7003; Fax: +34 95 213 7262.

E-mail addresses: mluque@uma.es; luisalej90@gmail.com.

Education is a key factor to explain economic growth (Barro 2001). Students' performance at early stages[2], particularly primary education, is clearly correlated with the national development of human capital accumulation and any factor affecting the quality of human capital production will directly affect future economic growth. Equally important is the role of education as an engine for promoting equity, to the extent that the educational system might reduce the impact of students' background on their academic results, as argued by the literature on equality of educational opportunities (Roemer 1998) and, hence, this contributes to higher levels of social mobility (Corak 2013; Galindo-Rueda and Vignoles 2005; Marcenaro et al. 2015). Education is supposed to provide students with the ability of progressing in their formation, especially in the case of the more disadvantaged ones which "(…) are, on average, more than twice as likely as students who are not considered disadvantaged to score in the bottom quarter of the performance distribution" (OECD 2013a p. 38). Those are crucial reasons to explain the growing number of empirical studies that have investigated the factors influencing students' performance.

The main emphasis of the previous literature on education production functions is on the relationship between a set of "quantitative" inputs and pupil outcomes; the latter measured –mainly– by some standardized scores. This means that other significant dimensions of the output have received less attention, in spite of the potential implications in terms of educational policies. For example, many of these papers pay little, if any, attention to the role of teachers' quality in the framework of the education production functions, which is an important limitation. Nevertheless this seems to be shifting to some extent, as the satisfaction of the teachers is a matter of increasing concern for the education policy; the recent Spanish Education Acts (LOE[3] –*Ley Órganica de Educación*– and LOMCE[4] –*Ley Orgánica para la Mejora de la Calidad Educativa*–) are a good example, insofar as they highlight the importance of improving teachers' working conditions and to incentive their consideration and social recognition.

In fact, it is often argued that, as reported by Barber and Mourshed (2007 p. 16), "the quality of an education system cannot exceed the quality of its teachers" and the esteem that teachers have in different societies varies widely, which means that teacher quality and its relationship with pupil performance is at the very heart of the debate about educational policy (Dolton and Marcenaro 2011). In other words, a good education system requires good teachers and, to achieve this aim, not only teachers with effective teaching skills are needed, but also with high levels of satisfaction. In this sense, we need to draw specific consideration to teachers and examine the extent to which variations in the way teachers feel in different schools may be a good proxy for the educational outcomes of primary school pupils. Additionally we should pay attention to the balance between that "qualitative" outcome and more standard measures of the output to the extent that, due to the conflict degree among these objectives, it could be impossible to find a feasible solution to simultaneously maximize all these outputs. Indeed, the pay-off matrix resulting from the multiobjective approach we follow (see section 5) has allowed us to establish the existence of conflicts among the objectives and, thus, it has shown that the Pareto optimal set is not limited to one solution.

---

[2] Pupils' test scores should be considered a good proxy of individual's ability (De Coulon et al. 2011).

[3] BOE 4th May 2006, Organic Law 2/2006.

[4] BOE 10th December 2013, Organic Law 8/2013.

Turning this logic around we suggest that the quality of teachers is likely to be higher if they are satisfied with their job and that teacher satisfaction at present is the main mechanism to attract and retain –in the long run– the best graduates into the education system. As the potential supply of teachers rises because of the higher satisfaction, entry into teaching will become more competitive. This, in turn, will mean that the average ability of those entering the job will rise. Consequently, we must determine not only the factors helping to achieve higher students' performance, but also higher levels of teacher satisfaction. In other words, we have to find a balance among those "outcomes" to "guarantee" the improvement of the education system as a whole.

Thus, our standpoint is that teacher satisfaction is an indicator of how good is the education system and, consequently, we need to account for this when evaluating the factors behind the improvement of the education system. So, our thesis is that policy makers should define increasing levels of teacher's satisfaction as an educational target at the same time that they seek to achieve higher levels of pupils' performance and lower proportions of students below the basic achievement level[5]. The latter is particularly relevant in terms of socioeconomic equity, because promoting lower proportions of students performing below a basic threshold will contribute to increase the opportunities of resilience. Espinoza (2007) highlights the importance of dealing with equity in a society that usually puts more effort on efficiency in resource management than in social justice. In the same vein, Schleicher (2014) relates equity to inclusion, to the extent that assuring equity by a minimum level of achievement may be a way of fostering social mobility and helping those students with disadvantages to overcome the difficulties that they are facing due to their characteristics.

Ideally, in order to define the best educational policy practices, we would like to know the extent to which the different proposed targets are attainable simultaneously or, by contrast, there are trade-offs. To achieve this we need to go a step further by combining traditional econometric estimates with multiobjective programming. The reason to choose the latter methodology is the existence of "bad outputs", which should be minimized, and "good outputs", which should be maximized throughout the teaching-learning process; the latter group includes qualitative outputs, such as teachers' satisfaction with their work. We intend to achieve these maximizing-minimizing objectives simultaneously.

Consequently, we seek to establish whether some factors can be affected to achieve a balanced Pareto optimal solution between an indicator of teachers' satisfaction and two different measures of students' performance (i.e. average students' performance and percentage of students achieving basic standards of learning). In order to measure the satisfaction of teachers, the teachers' career satisfaction index provided by the program on Trends in International Mathematics and Science Study (TIMSS) 2011, for Spain, has been used. This index is based on fourth grade teachers' reports on their agreement with some statements about their career's satisfaction.

[5] TIMSS 2011 mathematics assessments at the fourth grade developed a benchmarking scale, setting the basic level cut-off in 400 points with the intention of being as close as possible to the percentile scale used in TIMSS 1999 for eighth grade (396 points in mathematics and 410 in science) (Mullis et al. 2008). Students are grouped in each benchmark according to their scores. Each level also gathered which content and cognitive skills possessed the students which were classified in it. The basic level is defined as: "Students have some basic mathematical knowledge. Students can add and subtract whole numbers. They have some recognition of parallel and perpendicular lines, familiar geometric shapes, and coordinate maps. They can read and complete simple bar graphs and tables" (Mullis et al. 2012 p. 87).

By using fourth grade students[6] we ensure that the pupils enrolled in each class are taught by the teacher under scrutiny.

The interest in the analysis of the Spanish educational system can be seen in the results obtained in external evaluations as PISA or TIMSS, in comparison with other countries. In the case of PISA 2012 (OECD 2013b), Spain obtained scores in mathematics (484), reading (484) and sciences (496) very close to the average of the OECD in these competences (494, 496 and 501, respectively). Similarly, for TIMSS 2011 in fourth grade, Spain obtained 482 points in mathematics and 505 in Science, just around the average –500– for the countries participating in TIMSS (Martin et al. 2012 and Mullis et al. 2012). For that reason the analyses of the Spanish case might help to understand the performance of the educational systems in a broad perspective.

However, the distance respect to top performer countries (e.g. Finland, Republic of Korea or Singapore) is very large, reaching about two standard deviations. This has motivated policy makers to undergone many reforms of the Spanish educational system during the last decade. Namely, LOE was implemented in 2006 and was followed –recently– by the LOMCE in 2013, which is a modification of the previous one. Both acts paid more attention to the role of teachers; the LOE valued the importance of teachers by assuring that they receive a treatment and consideration in accordance with the relevance of their work. Thus, it intended to improve their working conditions and incentive their consideration and social recognition. Later, the LOMCE intended to provide teachers with more autonomy, in order to get specialized teachers and a more personalized education. It highlighted the importance of fostering ICT culture among teachers and it established the selection of head teachers by a competitive procedure based on the merits of teachers at the school.

The rest of the paper is organized as follows. The next section describes briefly the previous literature on this topic and, particularly, those papers focused on teachers' satisfaction. Section 3 describes our data. The methodological approach adopted and the main results of the empirical analysis are shown in sections 4 and 5. Section 6 concludes discussing some policy implications of our results.

2. Literature Review

Despite the potential importance of the link between good teachers and effective student learning, to the best of our knowledge, the literature on this is scarce in countries like Spain, mainly because of the difficulty to get access to good observational data to build a reliable indicator measuring "how good a teacher is". As a consequence, the Spanish education policy could not be well oriented in that subject, so it is not sensitive enough to substantiate the link between quality and effectiveness of teachers. In what follows we briefly review the literature on this topic to illuminate this less explored strand of the ongoing educational debate, paying special attention to papers exploring teachers' satisfaction as a proxy for teachers' quality. In this vein, Kumar (2014) found a positive correlation between teaching effectiveness at primary education level (the one we focus on) and job satisfaction. This means that teachers who were highly satisfied with their jobs were the most effective in teaching. Other interesting result of this study was the gender issue, as female teachers were more satisfied towards their job and a lot more effective in teaching than their male counterparts. This could be of relevance also in our model, so we have controlled for teacher gender in our estimates.

Also relevant is the paper by Michaelowa (2002), who studied primary teachers in Africa and found that there is a positive effect of teachers' job satisfaction on the quality of the education they are providing. However,

[6] The average age at time of testing was 9.8 years.

they remarked that there are many cases when this relationship does not happen, particularly with the implementation of measures based on more control to increase students' achievement –such as defining harder working conditions for teachers–, what reduces teachers' satisfaction. Other variables, e.g. classroom equipment, benefited teachers' performance and satisfaction. They also found that, in order to increase both teachers' satisfaction and students' achievement, teachers should not have qualifications higher than a Baccalaureate degree, because teachers with higher qualifications may face more difficulties when teaching in primary schools due to their higher aspirations.

Nonetheless this relation between teacher effectiveness and satisfaction is not limited to primary school, and it has been under evaluation more often at the secondary school level. Recently, Tek (2014) highlighted that those secondary schools which presented higher satisfied teachers performed better in mathematics. School leadership also played an important role in the interaction of satisfaction and achievement, to the extent that leaders can create a proper working climate where teachers can feel safe to try out new teaching methods. Similarly, Akiri (2014) analyzed some factors affecting teachers' career satisfaction in secondary school; he found that female teachers were more satisfied than males and that more qualified and experienced teachers were less satisfied, due to their feeling that they are not being rewarded enough –as previously stated by Michaelowa (2002)–. In addition, she declares that better salaries and working conditions might increase their satisfaction, what will help to retain highly qualified teachers.

Although the literature supports that the main determinants of teachers' satisfaction are factors as administrative support, student behaviors, workplace atmosphere, autonomy, efficacy and parental and community support, Patrick (2007) found that they did not explained completely teachers' satisfaction, due to the complexity of the definition of this concept.

In the field of self-assessment and motivation, Caprara et al. (2006) examined the contribution of teachers' self-efficacy beliefs and satisfaction at the school level, finding that teachers with a strong sense of efficacy had high levels of planning and organization. Likewise, Allinder (1994) highlighted this and also that teachers with these characteristics were open to experiment with new ideas and methods, in order to better meet the needs of their students (Cousins and Walker 1995). They also showed enthusiasm for teaching (Allinder 1994), what can have a positive influence on students' achievements and their own sense of efficacy (Podell and Soodak 1993; Tschannen-Moran and Woolfolk 2001).

In this vein, Klassen and Chiu (2010) found that teachers' years of experience exhibited nonlinear relationships with self-efficacy factors, which increased from the first steps of their careers until reaching the half of it, to fall afterwards. They also established that female teachers had a lower self-efficacy when managing classrooms and higher stress because of their amount of work and students' behaviours. Also in relation to the effect of teachers' gender and experience, Broughman (1997) –as Kumar (2014)– concluded that female teachers tend to be more satisfied than male school teachers, while teachers with less experience tend to be more satisfied than teachers with more experience. After controlling for other factors, his multivariate analysis indicated that salary did contribute to teachers' satisfaction in a positive way. But, although it was statistically significant, salary factors did not contribute so much to the prediction of teachers' satisfaction, what suggests that the satisfaction of teaching as a career has a higher effect on teachers' satisfaction than lucrative based satisfaction. In the case of the Spanish education system, public school teachers are paid very similarly, thus teachers' salary is not expected to account for variations in teacher effectiveness. However, to the extent that the variations in teachers' salary depend, mainly, on their years of experience, the latter could pick up partly the effect of the salaries.

Regarding with the alternative ways of measuring students' performance employed in this study, the high relevance of achieving equity in education –by means of assuring that students perform over a certain threshold– has been stressed as a relevant aim by the OECD, concretely on the second principle stated in the European Group for Research on Equity in Educational Systems (2005): "No student may leave the educative system while being below the minimum skill threshold, in order to have a decent life in the modern society". Demeuse and Baye (2008) also denoted that the OECD framework has indicated the importance of reducing disparities due to the likelihood of certain groups to perform in a low and unacceptable academic level. Other authors as Field et al. (2007) have stated that there is a high need of identifying those students who are underperforming and put a special effort on them. The importance of this output measure has been also emphasized by other studies as Choi and Calero (2013), who employed the lowest level of proficiency in PISA 2009 as dependent variable, highlighting its importance as an indicator of the risk of school failure; they also denoted that the increase in the risk of school failure was closely related to the reduction of education systems' equity. Similarly, Giménez et al. (2014) –who measured the performance change in the educational systems of 28 countries participating in TIMSS 2007 and 2011–, minimized the percentage of students under the basic threshold at the same time as they maximized students' achievement in mathematics in each country. One interesting result was that an orientation to good output (higher mean scores), or to both good and bad outputs, reduced the achievement of the education systems, while an orientation to the bad output (proportion of students below the basic level) increased the achievement of a country, and also the dispersion was lower. This good and bad outputs approach, which is going to be employed in the present study, has been commonly used in the field of efficiency in energy production, with the aim of studying the efficiency of firms' productive processes and their environmental repercussions[7]. However, to the best of our knowledge, it has received very little attention in the field of Economics of Education.

3. Data

The information analyzed in this paper comes largely from the latest release of TIMSS (2011). This program sampled students in a two-stage clustered sampling design, in which schools were selected in the first stage, and two math classes were randomly sampled within each of these schools in the second stage[8]. The main sample was designed to be nationally representative of pupils in the target age group and so the sampling criteria ("stratifiers") for each country are designed to address key characteristics of the nation's school system. The Spanish sample contains 151 schools, 200 teachers and 4,183 students; each student was univocally matched to one teacher. The use of the satisfaction index reduces our sample to 190 teachers, because there were 10 teachers who did not answer those questions related to the composition of this index. Within each sampled class, in principle, all students participated in the assessment. However, the reality is that the number of students included in the sample will be smaller than the class size because of students who did not participated (Williams et al. 2009). To account for this complex sampling design, sampling weights were applied and a jackknife resampling technique was employed to calculate standard errors correctly in statistical analysis.

[7] As, for example, Färe et al. (2005); Watanabe and Tanaka (2007); Sueyoshi and Goto (2011).

[8] Within each class in the sample, a minimum of 85 per cent of pupils were expected to participate.

As above mentioned, math teachers' career satisfaction index provided by TIMSS 2011 is based on fourth grade teachers' reports on their agreement with some statements about their satisfaction with their career: "I am content with my profession as a teacher"; "I am satisfied with being a teacher at this school"; "I had more enthusiasm when I began teaching than I have now"; "I do important work as a teacher"; "I plan to continue as a teacher for as long as I can"; and "I am frustrated as a teacher". Teachers' responses to those questions were averaged on a 4-point scale as follows: Agree a lot = 1; Agree a little = 2; Disagree a little = 3; and Disagree a lot = 4. After reverse-scoring relevant items, TIMSS 2011 assigned students whose teacher's average was 3 or greater to the Low level of the index, students with a teacher averaging between 2 and less than 3 were at the Medium level, and students with a teacher averaging less than 2 were assigned to the High level of the index. Students were scored according to their teachers' degree of agreement with the six statements on the scale. For example, students with "satisfied teachers" had a score of at least 10.1, due to their teachers "agreeing a lot" with three out of the six statements and "agreeing a little" with the rest, on average. Students with "less than satisfied teachers" had a score lower than 6.6, which is related to their teachers "disagreeing a little" with half of the six statements and "agreeing a little" with the rest, on average. The rest of the students had "somewhat satisfied teachers" (Mullis et al. 2012).

We seek to establish the contribution of some factors to achieve simultaneously a balance between that indicator of teachers' satisfaction and two different measures of students' performance in maths[9]: average students' performance by teacher and the percentage of students who do not achieve basic standards of learning[10] by teacher, i.e. with scores below 400 points. Basically, students below basic level are those who get scores below the average of the whole set of countries participating in TIMSS 2011 minus one standard deviation. This is related to international standards, so it is less related to a within country standard[11].

Full descriptive statistics for the sample under scrutiny are provided in Table A2 (Appendix A), reporting the mean and standard deviation for the exam scores in mathematics, for the proportion of pupils below the basic level as well as the teachers' satisfaction index for the whole set of variables being evaluated in the specifications reported[12]. In brief, Table A2 shows that, in classrooms where the proportion of students with 100 of more books at home is above 42.5%, the achievement both in terms of average scores and proportion of students above the basic level is much higher, and also the teachers' satisfaction index gets closer to the top. Similarly, when the

[9] For the sake of brevity and taking into account the high correlation between maths and sciences scores, we have focused on maths.

[10] Fifty countries participated in TIMSS 2011 in fourth grade for the international benchmark: Armenia; Australia; Austria; Azerbaijan; Bahrain; Belgium (Flemish); Chile; Chinese Taipei; Croatia; Czech Republic; Denmark; England; Finland; Georgia; Germany; Hong Kong; Hungary; Ireland; Islamic Republic of Iran; Italy; Japan; Kazakhstan; Kuwait; Lithuania; Malta; Morocco; Netherlands; New Zealand; Northern Ireland; Norway; Oman; Poland; Portugal; Qatar; Republic of Korea; Romania; Russian Federation; Saudi Arabia; Serbia; Singapore; Slovak Republic; Slovenia; Spain; Sweden; Thailand; Tunisia; Turkey; United Arab Emirates; United States; Yemen. In addition, there were seven additional benchmarking participants in fourth grade: Abu Dhabi (UAE); Alberta (Canada); Dubai (UAE); Florida (US); North Carolina (US); Ontario (Canada); Quebec (Canada).

[11] This is the international standard definition, which defines the following points as international benchmarks: advanced (625 points); high (550 points); intermediate (475 points); low (400 points) (Mullis et al. 2012 p. 87). The low level is also considered as basic.

[12] In Table A1 the labelling of the variables under scrutiny has been stated.

proportion of students who assert that their teacher is easy to understand is at the top, all the outcomes move towards increasing levels, although the trend is somehow attenuated for teachers' satisfaction. Particularly relevant for the latter outcome is the proportion of students who are congratulated by the teacher for their work in maths, however it seems not clearly correlated with the proportion below the basic performance level, which means that some trade-off is going on. By the same token, the pattern for the variable years of experience differs depending on whether we focus on teachers' satisfaction or mean scores in mathematics.

A stronger degree of conflict is stated for the variable accounting for the method used to evaluate the practice of teachers at school, because when the method is students' achievement the mean for teacher's satisfaction index is much lower than when the method is other, however the levels of students' achievement are higher.

Interestingly enough, these bivariate descriptive statistics allow to state that female teachers are more satisfied towards their job than their male counterparts, in line with the results reported by some of the authors highlighted in the literature review.

4. Econometric analysis

As aforementioned, we should not consider the success of an education system only by maximizing the average performance of the students, but also minimizing the percentage of students not achieving basic –international– standards of learning, as a way of promoting equity and, additionally, –from the teachers' side– maximizing teachers' satisfaction with their work, as the main mechanism to attract and retain the best graduates into the teaching profession.

Simple regression models have been estimated, in which the above mentioned dependent variables are regressed on the set of explanatory variables reported in section 3. Students' performance is a continuous standardized variable, the proportion of pupils below the basic performance level[13] is also continuous, but censored variable, and teachers' satisfaction index is continuous.

The coefficients obtained from estimating the regression model are used to build the three objective functions that we wish to optimize simultaneously (two of them to maximize and the other one to minimize), thus having a multiobjective programming problem.

As previously outlined, we can proxy the output of the educational process through different categories of students' performance. The level for each of these performance targets results from the combination of a set of individual and contextual features, unobservable factors and a random disturbance (ε). The idea behind the OLS estimator is to minimize the latter term in order to get rid of the so-called "statistical noise" as much as possible. If teachers' order number is indexed as *'i'*, and the three outputs considered (i.e. average students' performance for each teacher, proportion of students below basic level and teachers' satisfaction level) are indexed as *'j'*, this model can be represented by the following set of equations[14]:

$$P_j(i) = \hat{\alpha}^j + \hat{\beta}_1^j d1(i) + \hat{\beta}_2^j d2(i) + \cdots + \hat{\beta}_{19}^j saf(i) + \varepsilon_j(i) \qquad (1)$$

[13] To the extent that we are benchmarking against the full sample mean places Spain in a better relative position than benchmarking against, for example, the OECD or the EU means. From this point of view, we are providing a lower bound estimate of this indicator. Thus, the estimated coefficients for the second objective function have to be taken as a lower bound measurement of the relative indicator.

[14] This technique for modelling the educational process of exam performance is the educational production function.

$$i = 1, \ldots, n \ (number\ of\ observations, 190\ teachers);$$
$$j = 1,2,3 \ (1 = average\ scores\ in\ maths; \quad 2 = proportion\ below\ basic\ level;$$
$$3 = teachers'satisfaction).$$

where $P_j(i)$ is a measure of the output *'j'* for teacher *'i'*, and $d1(i), d2(i), \ldots, saf(i)$, a set of explanatory variables; $\varepsilon_j(i)$ is a random disturbance; $\hat{\beta}^j = \left(\hat{\beta}_1^j, \hat{\beta}_2^j, \ldots, \hat{\beta}_{19}^j\right)^T$ a vector of slope coefficients and $\hat{\alpha}^j$ a fixed but unknown population intercept. The size of the sample is represented by the integer value 'n'. Therefore, we are assuming that each of the three outputs is affected by random factors which are inherently unobservable and distributed normally. This parsimonious model is characterized by the parametric nature of its specification; the estimated parameters represent average educational production functions, i.e. the average combination of a set of educational inputs measured at teacher aggregated level transformed to obtain an average output level.

In order to get the best fitted model, many TIMSS 2011 variables for Spain were tested at teachers' level[15]. These variables were related to the level of teachers' qualifications, teachers' frequency of meeting with students' parents, mathematics teaching time per week, whether the use of calculators or computers was allowed or not during mathematic classes, frequency of homework assignment, attendance to professional development courses, the proportion of students who did like school, the proportion of students who enjoyed mathematics, the proportion of students whose parents checked their homework every day, other methods of teacher evaluation as observations by the principal (or senior staff), by inspectors or other persons external to the school and, lastly, teacher peer review and school problems as teachers arriving late at school or not attending.

TIMSS only differentiates between private and public schools, so it does not distinguish between private and semi-private schools[16]. The number of private schools for Spain is 49 and public schools are 102[17], but this variable was not significant when used in the model, so this distinction was not employed in the rest of the analysis.

An additional issue with regard to the regressors is that the variable which measures the level of safety which teachers feel in their school has been grouped into "agree a lot" or not. This is due to the lack of observations in the category "disagree a lot" and the presence of only 3% of observations in the category "disagree a little", what made us grouping these two mentioned categories with "agree a little". This let us differentiate among those who are sure about their feeling of high safety and those who are not (teachers who selected another option).

Table B1 (Appendix B) shows the estimated coefficients of the group of significant variables –in any of the three objective functions–; it also reports the standard deviations and the significance levels for each coefficient. The coefficients reported were normalized, although they are going to be interpreted relying on their original values in order to provide more intuitive explanations. As robustness check, given the censorship of the dependent variable in the case of the proportion of pupils below the basic performance level, we also estimated a Tobit model; we obtained very similar coefficients to those of OLS and slightly lower standard deviations, what gives further support to the robustness of our estimations. Alternatively, in order to account for the potential

[15] Only the significant ones –in any of the objective functions– were included in the final specification.

[16] Semi-private schools in Spain are private schools which are embraced in the regimen of legally established concert and are partially subsidized by the Government.

[17] In the final sample it is 49 and 95, respectively.

correlation among the error terms of the different equations we estimated a SURE model; again we obtained parameters of the same size and somewhat smaller standard deviations[18].

A set of variables was introduced with the aim of controlling the proportion of students –by teacher– who did not want to answer the corresponding questions, in the case of the students' questionnaire. Additionally, for the teachers' and school questionnaire we included missing flags representing which of these units did not answer the questions, respectively[19].

Focusing on the results, we find that teachers who have students with diglossia have lower average scores, being this negative effect increased as the proportion of students with diglossia grows. When the proportion is higher than 12%, the percentage of students under the basic level is increased by 4.2%[20]. A similar result was highlighted by Carabaña (2013), who found that diglossia decreased students' achievement and, thus, it increased the probability of grade repetition due, to some extent, to the correlation between diglossia and immigration status of the student, which is a matter of concern particularly in terms of equity. It could also be reflecting side effects of the involvement of co-official languages in the teaching-learning process.

The effect on students' math achievement of the proportion of students with 100 or more books in their home is particularly positive at the upper end of the percentage distribution. It also reduces the proportion of students under the basic level (in 8.3% for a teacher with 27.5% to 42.5% of his/her students with 100 or more books at home and in 7.5% when this percentage is higher than 42.5%). This variable was used as a proxy of cultural level by authors as Calero and Escardíbul (2013) and Carabaña (2013), which was said to increase students' performance. However, this variable does not exhibit a significant coefficient for the function explaining teachers' satisfaction, what reflects –somewhat– another source of simultaneity between student achievement and teacher satisfaction. The effect of the variable that controls for the proportion of students who did not answer the question reduces the average performance of teachers' students in 1.97 points and it would increase the proportion of students under the basic level in 0.75%, for each unitary percentage increase of the explanatory variable.

The same pattern as the variable of books at home can be found in the case of the proportion of students who declare that their teacher is easy to understand, which can be taken as a proxy for teachers' quality. Besides, it increases teachers' satisfaction by 0.898 index points in the case that the percentage of students declaring this ranges between 89% and 95.5%. Also affecting teachers' satisfaction, we found that those teachers who congratulate more than 76.5% of their students are 1.28 index points more satisfied than those in classrooms where the proportion was lower. However, neither the mean scores nor the percentage of students under the basic level are significant.

The students from schools whose head teacher agrees "a lot" that there is a great lack of mathematics' teachers in their school achieve, on average, 25.85 points less in mathematics, and those who only agree have less

---

[18] Tables are available upon request to the authors.

[19] Only missing flag variables with significant coefficients were reported in Table B1 (Appendix B) to conserve space.

[20] This figure has been obtained by undoing the normalization process of the pertinent normalized coefficient (0.281) in Table B1 (Appendix B). In order to do that, we have applied a proportional rule by using the differences between the maximum and minimum values –normalized and non-normalized, in Table A2, Appendix A– of the dependent variable "percentage of students under basic level" as follows: [(MaxNonNormalized-MinNonNormalized)*NormalizedCoefficient] /[(MaxNormalized-MinNormalized] = [(0.84-0)*0.28]/[4.87-(-0.73)]. The same rule has been employed for all the commented coefficients.

satisfied teachers (approximately 1 index point), probably as a consequence of the higher pressure that this lack of colleagues may generate on teachers' working conditions.

The evaluation method of teachers based on students' achievement has a positive effect on students' average scores and it also reduces the ratio of students who perform under the basic level in 6.3%. The missing flag for this variable shows that the uncertainty or confusion teachers could be facing about the control of this evaluation method could be in detriment of their satisfaction, as this method does not provide specific information for the improvement of teachers' capabilities (Peterson and Kauchak 1982).

Teachers' years of experience show an inverted U shape; the more the years of experience of the teacher, the higher the average performance of the students. Despite this positive effect, we found that it is marginally reduced when the number of years of experience increases, concretely, after 23.4 of years of experience. This variable was also analyzed by authors such as Goldhaber and Brewer (1997), who highlighted that students were benefited by the higher years of experience of their teachers. However –more recently– Decker et al. (2004) established that having no previous experience would be more beneficial and many other researchers suggest that teachers improve more at the beginning of their careers (Becker 1993; Rockoff 2004), although beyond their first few years, teachers may continue to improve their practices (Huberman 1992). This pattern reflects theories which state that less experienced teachers are often trying to survive in the classroom until they learn the curriculum, develop key classroom management skills and increase their instructional abilities (Johnson and Birkeland 2003). However, there is less agreement about the nature of returns to experience after these early years. Nevertheless, other research papers suggest that teachers who enjoy strong job protections may stop improving once they become established in their schools (Hansen 2009). Aside from this, to the extent that the effect of the missing flag variable on the mean scores in mathematics is positive and significant, we should consider the estimated coefficient as being downward biased.

An additional noteworthy result is that gender of the teacher does not seem to influence on any of the outcomes under evaluation. Thus, these potential correlations between both variables vanishes when conditioning on other factors.

Finally, the effect of a very high feeling of safety at the school by the teacher raises the average scores by 15.75 points and increases teachers' satisfaction, on average, by 1.39 index points, however it does not contribute to reduce the relative size of the group of disadvantaged students and, thus, it seems not to be contributing to higher levels of equity. This feeling of safety was studied by Gregory et al. (2012), who also found that this variable had a positive effect on teachers' satisfaction and performance.

The explanatory power of the estimated model is moderate[21], in line with those reported in the previous literature on Economics of Education. For example, MECD (2012)[22] examined the influence of socio-economic and cultural background on students' achievement, obtaining an explanatory power of 0.132 in reading, 0.168 in mathematics and 0.145 in science; Carabaña (2013) analyzed the likelihood of repetition of students in primary education and obtained a R^2 of 0.245 for standardized coefficients; Hanushek et al. (2014) studied the relationship between students' performance and parental and teachers' cognitive skills, together with country, school, parental

[21] R-squared of 0.466 for mean scores in mathematics, 0.375 for percentage of students under the basic level and 0.346 for teacher's satisfaction.

[22] Spanish Ministry of Education, Culture and Sports.

and students' characteristics, obtaining a R^2 of 0.260 for mathematics. Additionally, in the context of a mixed econometric multiobjective procedure –such as the one we have applied–, we find similar levels of variance explained by the models, e.g. in Wallenius et al. (1978) –who applied a multiobjective optimization methodology to study an econometric model of macroeconomics policy in Finland–, and –more recently– Marcenaro et al. (2010), who built a linear mixed integer multiobjective optimization problem from an econometric model that studied the satisfaction level of Spanish workers. Obviously, due to the limited explanatory power of most of the econometric models we should be cautious when interpreting the results as we cannot tell for sure that outputs would actually improve when other explanatory factors are at play. In other words, the potential improvement in all the three outputs is confined to the limited amount of variance explained by the models.

5. Specification and resolution of a Multiobjective Programming Model

5.1. Concepts and notation in Multiobjective Programming

In order to solve the multiobjective problem proposed in this study it is necessary to establish the basic definitions and notations. Considering the following general multiobjective problem:

$$\begin{aligned} &\max f(\mathbf{x}) = \big(f_1(\mathbf{x}), \dots, f_k(\mathbf{x})\big) \\ &s.t.: \qquad \mathbf{x} \in X \end{aligned} \tag{2}$$

involving k (≥ 2) conflicting objective functions $f_j\colon X \to \mathbb{R}$, which must be maximized simultaneously and where $\mathbf{x} = (x_1, \dots, x_m)^T$ are the decision variables. The decision vector $\mathbf{x}$ belongs to the feasible region $X \subset \mathbb{R}^n$, which is a nonempty compact set. The image of any $\mathbf{x} \in X$, $\mathbf{z} = f(\mathbf{x})$ is called objective vector and $Z = f(X)$ is called feasible objective region. Usually, it is impossible to find a feasible solution to simultaneously maximize all objective functions due to the conflict degree among the objectives. Because of that, the efficiency concept of optimality appears where none of the components can be improved without deteriorating at least one of the others: a decision vector $\boldsymbol{x}' \in X$ can be defined as efficient or Pareto optimal solution of the problem (2) if there does not exist another $\boldsymbol{x} \in X$ such that $f_j(\boldsymbol{x}') \leq f_j(\boldsymbol{x})$ for all $j = 1, \dots, k$, with at least one strict inequality. When this happens, $\boldsymbol{z}' = f(\boldsymbol{x}')$ is called nondominated objective vector. The efficient set is denoted by E and $f(E)$ is the nondominated objective set. A decision vector $\boldsymbol{x}' \in X$ is called weakly efficient or weakly Pareto optimal if there does not exist another $\boldsymbol{x} \in X$ such as $f_j(\boldsymbol{x}') < f_j(\boldsymbol{x})$ for all $j = 1, \dots, k$. The corresponding objective vector is called weakly nondominated objective vector. In addition, it is necessary to highlight that the set of efficient solutions is a subset of the weakly efficient solutions.

Furthermore, since the set of nondominated objective vectors contains more than one vector –usually many solutions–, it is useful to know the ranges of the objective vectors in the nondominated objective set. On the one hand, lower bounds are set by the nadir vector $\boldsymbol{z}^{nad} = \big(z_1^{nad}, \dots, z_k^{nad}\big)$, where $z_j^{nad} = \min_{\boldsymbol{x} \in E} f_j(\boldsymbol{x})$ for all $j = 1, \dots, k$, while upper bounds are given by the ideal values $\boldsymbol{z}^* = (z_1^*, \dots, z_k^*)$, where $z_j^* = \max_{\boldsymbol{x} \in E} f_j(\boldsymbol{x}) = \max_{\boldsymbol{x} \in X} f_j(\boldsymbol{x})$ for all $j = 1, \dots, k$. The nadir vector is not easy to obtain and when estimated from the pay-off table the values achieved are not necessarily good approximations (for details, see e.g. Ehrgott and Tenfelde-Podehl 2003; or Miettinen 1999). Recently, Deb et al. (2010) and Deb and Miettinen (2010) proposed more reliable approaches for its estimation. Both the ideal vector and the nadir vector are used frequently to normalize the objective functions since the range normalization is the most used one. In the case of this study, this normalization has been previously performed.

As all efficient solutions can be considered equivalent from the mathematical point of view, it is necessary to incorporate some preferential information into the model. Preferences about efficient solutions are commonly expressed by the so-called reference point $\boldsymbol{q} = (q_1, \dots, q_k)^T$**,** which consists of reference values for the objective functions. Given these values and a vector of weights $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$, the so-called achievement scalarizing function is built and minimized over the feasible set. Wierzbicki (1980) proposed one of the most commonly used achievement scalarizing functions:

$$s(\mathbf{q}, f(\mathbf{x}), \boldsymbol{\mu}) = \max_{j=1,\dots,k}\{\mu_j(q_j - f_j(\mathbf{x}))\} + \rho \textstyle\sum_{j=1}^{k} \mu_j(q_j - f_j(\mathbf{x})) \quad (3)$$

which must be minimized in the feasible region:

$$\begin{aligned} &\min \ s(\mathbf{q}, f(\mathbf{x}), \boldsymbol{\mu}) \\ &s.t.: \qquad \mathbf{x} \in X \end{aligned} \quad (4)$$

with $\mu_j > 0$ for all $j = 1, \dots, k$ and where $\rho > 0$ is a so-called augmentation coefficient, which must be a small value, and which assures the efficiency of the solutions generated (in many cases, ρ is equal to one thousandth, or one ten thousandth). Only the weak efficiency of the solution is assured when the second term is not used. The vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$ with $\mu_j > 0$ for all $j = 1, \dots, k$ is formed by the weights assigned to reach the reference values and can have different meanings (see Luque et al. 2009).

Nondominated solutions are produced by problem (4) and it is demonstrated that any Pareto optimal solution can be found by solving (4) using the ideal objective vector as reference point (or any objective vector that dominates it, as an utopian vector), and varying the weight vector in the whole weight vector space (Kaliszewski 1994). It is also demonstrated that any Pareto optimal solution can be found by solving (4) when fixing the weight vector and varying the reference point (Miettinen 1999). Recently, Luque et al. (2015) have demonstrated that, fixing the weight vector, it is possible to obtain a set of reference points which produces the same Pareto optimal solution.

A possible inconvenience of the achievement scalarizing function (3) to be solved in (4) is that it is generally non-differentiable, even if the objective functions of the problem (2) are all differentiable. However, this difficulty can be overcome by considering the following equivalent formulation to the problem (4):

$$\begin{aligned} &\min \quad \alpha + \rho \textstyle\sum_{j=1}^{k} \mu_j \left(q_j - f_j(\mathbf{x})\right) \\ &s.t.: \qquad \mu_j \left(q_j - f_j(\mathbf{x})\right) \le \alpha \qquad j = 1, \dots, k \\ &\qquad\qquad \mathbf{x} \in X \end{aligned} \quad (5)$$

Considering the differentiable formulation (5), the resulting single objective optimization problem will be solved using the optimization toolbox function for Matlab 2012[23].

5.2. Data characteristics

In our multiobjective model, let us consider all the variables described in the econometric study as decision variables, although not all of them are controllable by a decision maker, as we will discuss further in the Conclusions section. These variables can be seen in Table A1 (Appendix A).

5.3. Constraints

To make our multiobjective model more realistic, it is necessary to define a set of constraints. There is a set of technical constraints which ensure that certain binary variables do not take the value "1" simultaneously.

[23] Specifically, a function developed for Matlab Optimization Toolbox by Jonathan Currie was used (OPTI Toolbox), designed to solve single objective mixed-integer nonlinear programming problems.

Let us recall that the reference value of each group (which is assumed to be equal to 1 if the rest are equal to 0) is not considered a variable, which is the reason why the constraints reported in Table 5.1 are inequalities, except for the constraint (C7), which represents a quadratic constraint.

-Insert Table 5.1 here-

The restriction (C7) is the result of an inverted U shape relationship in the behaviour of teachers' years of experience, as it was previously mentioned in the econometric analysis (section 4). Thus, the problem becomes nonlinear, concretely quadratic, and then we have a mixed-integer nonlinear multiobjective optimization problem, also called mixed-integer quadratic multiobjective optimization problem.

In addition, other constraints have been derived from clear dependencies observed in the regression analysis. This means that we have chosen those pairs of variables whose dependencies were stronger according to this analysis, and thus it is not realistic to give them independent values. One of them is the variable '$yexp$'; all the dependencies between this variable and the rest included in our model have been checked, from where two-sided constraints have been build, using the 99% confidence intervals; these bounds for c and d are presented in Table 5.2. In this table the first row corresponds to the lower bound of c and d and the second row corresponds to their upper bound –for each variable–. In order to make this procedure more understandable, an example using the proportion of students taught by the teacher with 100 or more books at home, when this proportion is between 27.5% and 42.5%, is provided:

Dependency between '$yexp$' and '$b2$' is given by the following linear regression:

$$yexp = c \cdot b2 + d$$

where the confidence intervals of the coefficients are (at 99%):

$$c \in [c^l, c^u] = [-0.948, 8.762] \text{ and } d \in [d^l, d^u] = [17.075, 21.747]$$

which implies:

$$c^l \cdot b2 + d^l \leq yexp \leq c^u \cdot b2 + d^u$$

These inequalities are incorporated into the model by means of two new constraints:

$$yexp - (8.762 \cdot b2 + 21.747) \leq 0 \quad \text{(C10)}$$

$$yexp - (-0.948 \cdot b2 + 17.075) \geq 0 \quad \text{(C11)}$$

The same procedure will be applied in order to get the (C12) and (C13) constraints.

-Insert Table 5.2 here-

5.4. Objective functions

The objectives to be considered in this study are maximizing students' performance in mathematics, minimizing the percentage of students under the basic level and maximizing teachers' satisfaction index, all of them normalized. The econometric study has allowed us to express these outcomes as functions of a set of variables. Therefore, $\hat{\beta}_m^j$ is the regression coefficient of variable m for performance level j, and $\hat{\alpha}^j$ is the independent term of performance level j, then we have the following 3 objectives:

$$EP_j(\boldsymbol{x}) = \left(\hat{\beta}^j\right)^T . \mathbf{x} + \hat{\alpha}^j \quad (6)$$

$$j = 1,2,3.$$

where $\left(\hat{\beta}^j\right)^T = (\hat{\beta}_1^j, \hat{\beta}_2^j, \dots, \hat{\beta}_{19}^j)$ and $\boldsymbol{x} = (x_1, x_2, \dots, x_{19})^T$, which measure the expected performance in mathematics, the expected percentage of students under the basic level and the expected teachers' satisfaction index. The resulting multiobjective problem to be solved in each case is the following:

$$Max\ (EP_1(\boldsymbol{x}), -EP_2(\boldsymbol{x}),\ EP_3(\boldsymbol{x})) = \left(\left(\hat{\beta}^1\right)^T.\mathbf{x} + \hat{\alpha}^1, -\left(\hat{\beta}^2\right)^T.\mathbf{x} - \hat{\alpha}^2,\ \left(\hat{\beta}^3\right)^T.\mathbf{x} + \hat{\alpha}^3\right) \quad (7)$$

Subject to: (C1) – (C13)

Thus, the model under scrutiny is a mixed integer quadratic multiobjective model with 3 objectives, 19 variables and 13 technical constraints. To solve any multiobjective problem, it is usual to consider all the objective functions in the same sense (maximize or minimize) since that to minimize any objective function is equivalent to maximize the opposite one. In our case, we have changed the sign of the second objective function taking into account the formulation of the section 5.1 and thus, it is taken into account in the resolution of the problem (7).

For each case, we have calculated their ideal values:

$$z^* = (539.938, 0.029, 14.522)$$

These ideal values, in teacher level, are 539.938 points of students' average score, 2.9% of students performing under the basic level and a score of 14.522 in the teacher satisfaction index.

The following pay-off matrix of the problem (see below) proves that there is actually some degree of conflict between each pair of functions, this meaning that the use of a multiobjective approach is appropriate. The matrix displays the values of the three objective functions in each of the individual optima (for example, row 1 corresponds to the values that each of the objective functions –in columns– take for the optimal solution for f_1 and so on). The worst value of each column (491.283 for f_1, 0.198 for f_2 and 10.950 for f_3) indicates the corresponding anti-ideal value. These values can overestimate their corresponding nadir values and thus, the ranges of variation of the nondominated objective vectors can be even wider. For more details, see Miettinen (1999) and Deb et al. (2010).

Pay-off matrix (not normalized values)

$$\begin{array}{c} \\ x_1^* \rightarrow \\ x_2^* \rightarrow \\ x_3^* \rightarrow \end{array} \begin{array}{c} \begin{array}{ccc} f_1 & f_2 & f_3 \end{array} \\ \begin{bmatrix} 539.938 & 0.037 & 12.343 \\ 513.812 & 0.029 & 10.950 \\ 491.283 & 0.198 & 14.522 \end{bmatrix} \end{array}$$

5.5. Solutions of the Multiobjective models

As previously mentioned, the multiobjective problem has the aim of detecting the profile of the most balanced teacher in terms of these three selected outcomes. To this end we have used a reference point approach, where the reference value for each dependent variable has been obtained from the actual teachers who took the TIMSS survey in Spain. In particular, the related characteristics of the teacher who gets the best values for the three objectives is the following: the students of this teacher obtain an average score of 568.590, the proportion of students under the basic level is 0% and the value of the teacher satisfaction index is 13.390. Furthermore, using these three reference values which belong to an actual teacher makes this analysis more realistic. To sum up, the reference values are:

$$q_1 = 568.590,\ \ q_2 = 0,\ \ q_3 = 13.390$$

Besides, note that these objectives are not in the same scale and since that nadir values are not available, we decided to work with all the values normalized: reference values and linear coefficients of the objective functions –which are displayed in Table B1 (Appendix B) –. Thus, the normalized reference values are:

$$q_1 = 2.220,\ \ q_2 = -0.730,\ \ q_3 = 1.140$$

and consequently the problem (5) is given by:

$$Minimize \quad \alpha + \rho\left(\left(q_1 - \left(\hat{\beta}^1\right)^T.\mathbf{x} - \hat{\alpha}^1\right) + \left(\left(\hat{\beta}^2\right)^T.\mathbf{x} + \hat{\alpha}^2 - q_2\right) + \left(q_3 - \left(\hat{\beta}^3\right)^T.\mathbf{x} - \hat{\alpha}^3\right)\right)$$

$$Subject\ to: \ q_j - (\hat{\beta}^j)^T.\mathbf{x} - \hat{\alpha}^j \leq \alpha \qquad j = 1,\ 3 \tag{8}$$

$$(\hat{\beta}^2)^T.\mathbf{x} + \hat{\alpha}^2 - q_2 \leq \alpha$$

(C1) – (C13)

All the criteria are equally weighted, meaning that it is implicitly assumed that the achievement of all the reference levels have the same importance for the decision maker. The solutions obtained after solving the previous problem are shown in Table 5.3. Nevertheless, in section 5.6, we perform alternative analyses with different combinations of weights to check the robustness of the results.

-Insert Table 5.3 here-

The results presented in this table provide a "taxonomy" of the best and more balanced teachers, in terms of the average score of their students, proportion of students under basic level and teacher satisfaction index. Analyzing the solutions generated, it is remarkable that multiobjective programming is able to detect some issues that can hardly be obtained only with the econometric analysis. For example, if we only optimize the second objective function –percentage of students under the basic level– it would be better if the variable $b2$ was equal to 1 instead of $b3$, because the coefficient of $b2$ (-0.556) is lower than the corresponding of $b3$ (-0.5); however, this change in the solution deteriorates the value of the first objective function –the performance in mathematics–, since the coefficient of $b2$ for this function (0.599) is lower than the corresponding of $b3$ (0.778) and, as consequence, the optimal value of alpha in (5) would also deteriorate. This means that, when a multiobjective approach is used –in our case by means of solving the problem (8)–, it is better if $b2$ is equal to 0 and $b3$ equal to 1.

In summary, regarding to the results presented in Table 5.3, we could conclude that when teachers do not have students with diglossia they obtain better and balanced optimum results in the three dependent variables –what raises some social and cultural concerns that will be discussed in the conclusions section–. This also happens when the proportion of students with 100 or more books is higher than 42.5%, the proportion of students who say that their teacher is easy to understand is higher than 95.5% and the proportion of students who are congratulated by their teacher surpasses 76.5%. Thus, for these categories the upper bounds of the variables seem to work in the same direction. When the head teacher of the school where teachers work thinks that there is not a lack of teachers in his/her school –in mathematics– it has the same effect, or when the evaluation methods –for teachers– used at the school are based on students' achievement. Those teachers with a number of years of experience around the median value and/or who feel very safe at their school also contribute to a better level of balanced outcomes. Last, but not least, being a female teacher helps to achieve this balanced optimum.

5.6. Robustness of the solutions

In addition to the previous analysis, we have also evaluated the robustness of the solutions by changing the weights of the reference values and the confidence level for the constraint intervals. In the case of the weights, the solution did not change when each weight at a time was increased by one unit and when we assigned double weight to mean scores and percentage of students under basic level with respect to teachers' satisfaction, what is denoting that solutions are robust with respect to these parameters[24].

In the case of the confidence intervals for the dependence constraints of the pair of variables, they have been changed to a 95% confidence interval, but the values of the objective functions changed slightly: '$yexp$'

[24] For reasons of space the corresponding table is not presented, but it may be supplied upon request to the authors.

and ‘$yexp2$’ variables turned into 21.375 and 456.891 with respect to those in Table 5.3, so they were not very influenced by this variation. The biggest difference in the values of the objective functions with respect to the 99% confidence intervals is irrelevant (average scores in mathematics changed from 539.938 to 539.691, while the other two dependent variables did not change when considering three decimals, which is the criteria followed in this study), thus meaningless.

6. Conclusions

The literature on education production functions, in the case of Spain, has paid little attention to the role of teachers' quality. We suggest that the quality of teachers is likely to be higher if they are satisfied with their job and that teacher satisfaction at present is the main mechanism to attract and retain –in the medium and long run– the best graduates into the education system. In other words, we propose using teachers' satisfaction as an additional indicator –together with students' average scores– of how good is an education system.

Additionally, to the extent that assuring equity by a minimum level of achievement can be a way of fostering social mobility and helping those students with disadvantages to overcome the difficulties that they are facing due to their characteristics, we include the percentage of students under the basic threshold as a third output under analysis.

An econometric analysis has allowed us to determine the significant parameters which relate the outcomes to many inputs and, then, a multiobjective approach has contributed to the identification of the desirable profiles of teachers who reach good results in the three outcomes (math teachers' satisfaction, average scores in mathematics and the proportion of students who perform under the basic level in mathematics), as a source of information for determining policies that might be carried out in order to improve their performance. From a methodological point of view, we look for generating Pareto optimal solutions through the methodology of reference point that allows us to reach or get as close as possible to some reference values.

Thus, this study provides a contribution to the design of education policies, which is particularly interesting in a context of economic crisis when the budget assigned to education in Spain has been substantially cut. The results obtained can be translated into suggestions for teachers, on the one hand, and for educational authorities, on the other hand, because although all inputs might be considered as decision variables, not all of them are controllable by both groups of decision makers. In this sense, from our results it should be inferred that teachers should attend specific training courses to keep improving the way they communicate their knowledge to the students and also reinforce the interaction techniques to congratulate them as a feedback when they are performing correctly. These factors could also be promoted by the educational authorities reinforcing the curriculum offered by the Institutes of Education in terms of communication skills for future teachers. Likewise the lack of specialized teachers at schools to cover all lessons could be harmful to the achievement of these objectives, so it would be necessary to solve this problem before beginning with the course. This conclusion has very important implications in terms of education policy as far as one of the most controversial policy measures implemented by the Spanish education authorities, as a consequence of the economic crisis, has been to delay the replacement of teachers[25]; in fact very recently[26] the Spanish Government has announced that primary school teachers will be replaced from the moment they are unable to teach.

Teachers' experience is also an important issue; the econometric estimates show that teachers' tenure has significant impact on students' average scores, but not on teachers' satisfaction and the ratio of students who perform under the basic level. However, it contributes positively to the achievement of a balanced Pareto optimal solution among the three different outcomes, although till certain threshold; then, they could have a negative effect (inverted U shape). This inflexion point does not appear around the average years of reaching tenure (9.5)[27], thus we cannot link achieving a permanent position as being the starting point on the decrease in satisfaction by the teacher, what contradicts the argument that teachers who enjoy strong job protection may stop improving once they become established in their schools (Hansen 2009). Besides student performance is reduced from the moment teachers reach 23 years of experience, which could partly be a reflection of the reduction in teacher satisfaction a decade before. Additionally, to the extent that the variations in teachers' salary depend –mainly– on their years of experience, the decreasing marginal effect of the experience could be reflecting a lack of incentives to boost teaching efforts, resulting from a nearly flat career progression. This is a factor clearly under control of policy makers, as they could change this by creating career progression routes based on a balanced target for different education outputs (as the one presented here).

In addition, policy makers should be more concerned with the need to promote some cultural habits, concretely promoting reading habits by raising conscience to parents about the importance of books via, e.g., advertising campaigns. Teachers also play an important role in promoting reading habits among students and interacting with their parents on this issue. This recommendation is in line with the equity issue raised in the Introduction section of our research: it is necessary to develop strategies that enable relatively disadvantaged students to achieve basic standards. This threshold is an important indicator of the risk of school failure which correlates with higher unemployment rates and translate in high costs for the individual and the society as a whole. In fact, Dolado (2009) estimated that the direct and indirect cost rigged to the high rate of school failure in Spain is equivalent to 0.43% of the GDP.

Related to that we found that the percentage of students with diglossia is detrimental for the education performance –in terms of the three outputs under scrutiny–.This does not mean –of course– that speaking different languages is not positive, however policy makers should put greater effort on the integration of pupils from different cultural backgrounds (either immigrants and/or speakers of co-official languages), what will additionally contribute to social cohesion.

Finally, the feeling of safety of teachers in the field of their personal integrity and also in their job conditions has a very positive influence on the achievement of the objectives. Consequently creating a comfortable working environment should be a key principle in every school, what is something under control of education authorities and, particularly, head teachers. Thus, this gives some support to the LOMCE on the importance of the

[25] The education authorities do not provide figures on this, but many newspapers have published news highlighting this issue and the strikes undertaken in different autonomous communities to complain about it (see, e.g. http://www.diariodesevilla.es/article/sevilla/1495842/la/sustitucion/profesor/tarda/cinco/veces/mas/por/los/recortes.html, or http://ccaa.elpais.com/ccaa/2014/03/27/catalunya/1395933405_126552.html).

[26] See, e.g., http://www.elmundo.es/sociedad/2015/09/29/560a5e9346163f83418b457d.html.

[27] Authors own calculations based on TALIS 2013.

selection of head teachers by a competitive procedure, following the tradition of other developed countries (UK, Finland, etc.).

Interestingly enough the solutions of our multiobjective approach shows that teacher satisfaction is the only indicator whose optimal value exceeds the sample maximum, implying that it is where more effort needs to be placed and, at the same time, greater scope of intervention is possible. Obviously, as in any empirical implementation using real data, to the extent that our models account for less than a half of the variance of each dependent variable, we should be cautious when analyzing the implications of our results as there are environmental variables beyond educational authorities' control. Nevertheless, many other factors have been considered in the alternative estimated models which did not show any significant impact on the outcomes evaluated (e.g. teachers' frequency of meeting with students' parents, mathematics teaching time per week, whether the use of calculators or computers was allowed or not during mathematic classes, frequency of homework assignment, the proportion of students whose parents checked their homework every day, etc.); this provides some additional confidence on the importance of dealing with some of the factors mentioned in this concluding section.

We can conclude that, consistently with the challenges that educational national policies and practices are facing –as those that the most recent education acts (LOE and LOMCE) highlight–, improving teacher satisfaction appears to be a promising area for future research to identify how other countries have achieved both excellence and equity in students' achievement.

-Insert Appendix A here-

-Insert Appendix B here-

References

Akiri, A. A. (2014). Teachers' Career Satisfaction and Students' Academic Performance in Delta Public Secondary Schools. *Journal of Educational and Social Research, 4(1)*, 267–272.

Allinder, R. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, *17,* 86–95.

Barber, M., & Mourshed, M. (2007). *How the world's best performing school systems come out on top*. New York: McKinsey & Company.

Barro, R. (2001). Human capital and growth. *American Economic Review*, *91(2)*, 12–17.

Becker, G. (1993). *Human Capital* (3rd ed.). Chicago: University of Chicago Press.

Broughman, S. P. (1997). *Job Satisfaction Among America's Teachers: Effects of Workplace Conditions, Background Characteristics, and Teacher Compensation*. Washington, DC: National Center for Education Statistics, U.S. Department of Education, Office of Educational Research and Improvement.

Calero, J., & Escardíbul, O. (2013). El rendimiento del alumnado de origen inmigrante en PISA 2012. In INEE (Ed.), *PISA 2012: Programa para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario* (pp. 4–31). Madrid: Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44*, 473–490.

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.

Choi, A., & Calero, J. (2013). Determinants of the risk of school failure in Spain in PISA-2009 and proposals for reform. *Revista de Educación, 362*, 562–593.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives, 27(3)*, 79–102.

Cousins, J., & Walker, C. (1995). *Personal teacher efficacy as a predictor of teachers' attitudes toward applied educational research*. Paper presented at the annual meeting of the Canadian Association for the Study of Educational Administration, Montreal.

De Coulon, A., Marcenaro, O., & Vignoles, A. (2011). The value of basic skills in the British labour market. *Oxford Economic Papers, 63(1)*, 27–48.

Deb, K., & Miettinen, K. (2010). Nadir point estimation using evolutionary approaches: better accuracy and computational speed through focused search. In M. Ehrgott, B. Naujoks, T. J. Stewart and J. Wallenius (Eds.), *Multiple criteria decision making for sustainable energy and transportation systems* (pp. 339–354). Berlin: Springer-Verlag.

Deb, K., Miettinen, K., & Chaudhuri, S. (2010). Towards an estimation of nadir objective vector using a hybrid of evolutionary and local search approaches. *IEEE Transactions on Evolutionary Computation*, *14(6),* 821–841.

Decker, P., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach for America: Findings from a national evaluation*. Princeton, New York: Mathematica Policy Research.

Demeuse, M., & Baye, A. (2008). Measuring and comparing equity in education systems in Europe. In N. Sogeul and P. Jacard (Eds.), *Governance and performance of education systems* (pp. 131–142). Dordrecht, the Netherlands: Springer.

Dolado, J. J. (2009). La educación en la encrucijada. In FEDEA (Ed.), *La Crisis de la Economía Española. Lecciones y Propuestas*. (pp. 56–58). Madrid: FEDEA, Sociedad abierta.

Dolton, P., & Marcenaro, O. D. (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.

Ehrgott, M., & Tenfelde-Podehl, D. (2003). Computation of ideal and nadir values and implications for their use in MCDM methods. *European Journal of Operational Research*, *151*, 119–139.

Espinoza, O. (2007). Solving the Equity–Equality Conceptual Dilemma: A New Model for Analysis of the Educational Process. *Educational Research, 49(4)*, 343–363.

European Group for Research on Equity in Educational Systems (2005). *Equity in European educational systems: A set of indicators*. Liege, Belgium: European Commission Directorate General of Education and Culture.

Färe, R., Grosskopf, S., Noh, D. W., & Weber, W. W. (2005). Characteristics of a polluting technology: theory and practice. *Journal of Econometrics, 126(2)*, 469–492.

Field, S., Kuczera, M., & Pont, B. (2007). *No more failures: Ten steps to equity in education*. Paris, France: OECD.

Galindo-Rueda, F., & Vignoles, A. (2005). The Declining Relative Importance of Ability in Predicting Educational Attainment. *The Journal of Human Resources, 40(2)*, 335–353.

Giménez, V., Thieme, C., Tortosa-Ausina, E., & Prior, D. (2014). *Una comparación internacional de sistemas educativos: una aplicación del índice global de Malmquist-Luenberger*. Paper presented at XVII Encuentro de Economía Aplicada, Gran Canaria.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources, 32(3)*, 505–523.

Gregory, A., Cornell, D., & Fan, X. (2012). Teacher Safety and Authoritative School Climate in High Schools. *American Journal of Education, 118(4)*, 401–425.

Hansen, M. (2009). How career concerns influence public workers' effort: Evidence from the teacher labor market. *Urban Institute Working Paper, No. 40*.

Hanushek, E., Piopiunik, M., & Wiederhold, S. (2014). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. *NBER working papers series, No. 20727*.

Huberman, M. (1992). Teacher development and instructional mastery. In A. Hargreaves and M. G. Fullan (Eds.), *Understand Teacher Development* (pp. 122–142). New York: Teachers College Press.

Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a 'sense of success': New teachers explain their career decisions. *American Educational Research Journal, 40(3)*, 581–617.

Kaliszewski, I. (1994). *Quantitative Pareto Analysis by Cone Separation Technique*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Klassen, R. M., & 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.

Kumar, P. (2014). A study of job satisfaction of British education trained teacher working at primary schools in Saharanpur district. *Journal of Indian Research, 2*, 141–145.

Luque, M., Lopez-Agudo, L. A., & Marcenaro, O. D. (2015). Equivalent reference points in multiobjective programming. *Expert Systems with Applications, 42(4)*, 2205–2212.

Luque, M., Miettinen, K., Eskelinen, P., & Ruiz, F. (2009). Incorporating Preference Information in Interactive Reference Point Methods for Multiobjective Optimization. *OMEGA-International Journal of Management Science, 37(2),* 450–462.

Marcenaro, O. D., Luque, M., & Ruiz, F. (2010). An application of multiobjective programming to the study of workers' satisfaction in the Spanish labour market. *European Journal of Operational Research*, 203, 430–443.

Marcenaro, O. D., Micklewright, J., & Vignoles, A. (2015). Social mobility, parental help, and the importance of networks: evidence for Britain. *Longitudinal and life course studies*, *6(2)*, 190–211.

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

MECD (2012). *PIRLS - TIMSS 2011 Estudio Internacional de progreso en comprensión lectora, matemáticas y ciencias IEA Volumen I: Informe Español*. Madrid: Instituto Nacional de Evaluación Educativa.

Michaelowa, K. (2002). Teacher job satisfaction, student achievement, and the cost of primary education in Francophone Sub-Saharan Africa. *HWWA Discussion Paper, No. 188.*

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic Publishers.

Mullis, I. V. S., Erberber, E., & Preuschof, C. (2008). The TIMSS 2007 International Benchmarks of Student Achievement in Mathematics and Science. In J. F. Olson, M. O. Martin and I. V. S. Mullis (Eds.), *TIMSS 2007 Technical Report* (pp. 339–347). Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center, Boston College.

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.

OECD (2013a). *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II)*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201132-en

OECD (2013b). *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know.* PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264208780-en

Patrick, A. S. (2007). *Examination of Teacher Workplace Satisfaction and Student Achievement*. PhD Dissertation at the Georgia Southern University.

Peterson, K., & Kauchak, D. P. (1982). *Teacher evaluation: Perspectives, practices, and promises*. Salt Lake City: Graduate School of Education, University of Utah.

Podell, D., & Soodak, L. (1993). Teacher efficacy and bias in special education referrals. *Journal of Educational Research*, *86,* 247−253.

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

Roemer, J. E. (1998). *Equality of opportunity*. Cambridge, Massachusetts: Harvard University Press.

Schleicher, A. (2014). Charting a way towards equity and excellence. In OECD (Ed.), *Equity, excellence and inclusiveness in education – policy lessons from around the world* (pp. 13–22). Paris: OECD Publisher.

Sueyoshi, T., & Goto, M. (2011). Measurement of returns to scale and damages to scale for DEA-based operational and environmental assessment: How to manage desirable (good) and undesirable (bad) outputs? *European Journal of Operational Research*, *211(1),* 76–89.

Tek, B. A. (2014). *An investigation of the relationship between school leadership, teacher job satisfaction, and student achievement*. PhD Dissertation in Education at the University of Rhode Island.

Tschannen-Moran, M., & Woolfolk, A. (2001). Teacher efficacy: Capturing an elusive concept. *Teaching and Teacher Education*, *17*, 783−805.

Wallenius, H., Wallenius, J., & Vartia, P. (1978). An Approach to Solving Multiple Criteria Macro-Economic Policy Problems and an Application. *Management Science*, *24(10)*, 1021–1030.

Watanabe, M., & Tanaka, K. (2007). Efficiency analysis of Chinese industry: A directional distance function approach. *Energy Policy, 35(12)*, 6323–6331.

Wierzbicki, A. P. (1980). The use of reference objectives in multiobjective optimization. In G. Fandel and T. Gal (Eds.), *Multiple Criteria Decision Making, Theory and Applications* (pp. 468–486). Berlin: Springer-Verlag.

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., Smith, C., & Stearns, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide*. Washington, DC: National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences.