Evaluating the Global Efficiency of Teachers through a Multi-criteria Approach

Mariano Luque[a], Oscar D. Marcenaro-Gutierrez[b], Ana B. Ruiz[a,*]

[a] *Department of Applied Economics (Mathematics)*
[b] *Department of Applied Economics (Statistics and Econometrics)*
Universidad de Málaga, C/ Ejido, 6, 29071, Málaga, Spain
mluque@uma.es, odmarcenaro@uma.es, abruiz@uma.es

Abstract

Improving the quality of education is one of the main objectives for governments, given that it plays an essential role in providing well-being to the population. In this regard, the efficiency of teachers may be a key issue to improve the students' academic performance and learning achievement. Specifically, in this paper, we analyse the efficiency of teachers from a multi-criteria optimization perspective. The purpose is to detect the most efficient teachers to gain knowledge about the characteristics of their teaching context, which contribute to increase the education performance of the students. To this aim, data from TIMSS and PIRLS 2011 for fourth-grade reading, mathematics and science teachers in Spain have been used. Three synthetic indexes gathering these data are calculated: a weak index allowing full compensation among the variables, a strong index informing about the worst variable value (no compensation scheme) and a synthetic index representing a different degree of compensation of the two formers. These indexes are built using aspiration and reservation levels for each variable. This allows us to include desirable ranges of values to be achieved by the variables in the study of the teachers' efficiency. With this, it is possible to know if the teachers are performing better, within or worse than the desired limits, and to rank the teachers according to their efficiency. The analysis carried out has contributed to have a better understanding of the policies to be followed in order to promote efficiency in teaching.

Keywords: efficiency; teachers; synthetic index; primary education.

---

[*]Corresponding author. Permanent address: Department of Applied Economics (Mathematics), Universidad de Málaga, C/ Ejido, 6, 29071, Málaga, Spain. E-mail address: abruiz@uma.es. Tel.: +34 952131156. Fax: +34 952132061.

1. Introduction

Improving the quality of education is one of the main objectives for governments, given that education has a direct effect on the social capital, economic growth and well-being of a country [1]. Indeed, the quality of education is so closely related to the development of population's cognitive skills that authors as [1] used them as its proxy. In this regard, researchers and practitioners must put special attention to the study of the "quality" of teachers, since having effective and efficient teachers may be a key issue to improve students' academic performance and learning achievement. By making an efficient use of the educational resources available, teachers could be able to increase their students' performance and raise the implication of the latter in the learning process. But, how to enhance the efficiency of teachers? The answer to this question requires, previously, identifying the main features of these teachers regarded as efficient. This is the main focus of this paper, which –based on real data analyses– intends to provide a new methodological approach to determine the features of the most efficient teachers.

Traditionally an efficient teacher has been considered as the one that is able to achieve the maximum outputs from a given amount of inputs of the teaching-learning process, or in the other way around, as the teacher capable to reach the same outputs with the minimum quantity of inputs [2]. In this context, this study aims to measure teachers' efficiency and identify the characteristics of the most efficient teachers, as a way to detect effectiveness. The efficiency has been measured as a proxy of not only the students' performance (output variable), but also using an engagement index indicating the capacity of teachers to engage their students (output variable) and taking into account several variables regarding the characteristics of the learning environment (input variables). To this end, the PIRLS (Progress in International Reading Literacy Study) and TIMSS (Trends in Mathematics and Science Study) databases have been combined in their 2011 wave for fourth-grade (9/10 years old) Spanish students. Fourth-grade students were chosen, as they are more malleable at this age [3] and would thus better reflect their teachers' procedures. Actually, Spanish primary school students have the same teacher in each cycle. Hence, a teacher in third- and fourth-grade teaches the same students, so their engagement could be attributed to the teachers under scrutiny. In addition, both databases contain a large number of teacher-level variables gathered in their teacher questionnaire –directly linkable to the students they taught– and a measurement of the students' engagement, which are not included in other international assessments such as, for example, PISA (Programme for International Student Assessment).

Taking into account the idea of optimizing input and output variables simultaneously to measure teachers' efficiency, in this work, we propose to study this efficiency from a wider multi-criteria optimization perspective. We suggest measuring the efficiency considering both inputs and outputs with the same importance, assuming that efficient teachers are those achieving, at the same time, optimal values for all the inputs and outputs. Given that the maximization of outputs at the minimum inputs represent a conflict, the multi-criteria optimization nature of the problem under study is unquestionable. With this, we want to have a better insight about how to distribute the available resources to improve key factors of the teaching-learning process that are desirable to let teachers achieve efficiency.

As the study we are dealing with, many real-life problems implies the simultaneous optimization of several objectives or criteria, which are in conflict. Such problems are called either multi-objective optimization problems if the set of feasible solutions is infinite and not explicitly known, but defined by a set of constraints (continuous case), or multi-criteria optimization problems if there is a finite set of feasible alternative or solutions (discrete case). The conflicting nature of the objectives makes it impossible to find a single solution to optimize all of them at the same time. Instead, there exists a set of Pareto optimal or efficient solutions at which no objective value can be improved without worsening, at least, one of the others.

In general, a common way to evaluate efficiency in any field consists of building synthetic indicators or indexes, which can indeed be calculated based on multi-criteria optimization approaches. In this paper, we build a set of synthetic indexes that are capable of measuring the efficiency of the teachers in our sample (the finite set of feasible alternatives, in our case) under the multi-criteria optimization perspective we have mentioned, using the input and output variables assessing the different aspects of teachers' performance at the same time. These synthetic indexes take into account different aggregation schemes that allow different grades of compensation among the criteria, varying from the total compensation to no compensation. This enables us to rank the teachers according to their efficiency level in different ways, with the purpose of analysing the main features and the learning context of the most efficient ones.

The synthetic indicators considered are built using the double reference point (DRP) approach [4], a multi-criteria optimization technique where reference values that are desirable and acceptable for each of the criteria are considered, called aspiration and reservation levels, respectively. With this, we introduce into the indexes' building process the use of a threshold with desirable and acceptable levels to be achieved by each variable, enabling the efficiency of each teacher to be evaluated depending on its performance regarding these values. All the values assigned by the DRP and, thus, the indexes are measured in the same scales. Therefore, they can be easily interpreted and provide information of the teachers' efficiency from a global point of view, and not only individually for each variable, using different degrees of compensation among the variables. Previously, this methodology has been used in other fields, which will be cited in the next section. However, to the best of our knowledge, it is the first time used to evaluate the efficiency of teachers.

The rest of the paper is organised as follows. Section 2 summarises the literature in this field. Section 3 describes the data used in our analysis to evaluate and measure the teachers' efficiency, providing descriptive statistics of the sample under scrutiny. Section 4 states some multi-criteria optimization concepts and the methodological process to build the synthetic indexes. The results obtained are discussed in Section 5. Finally, concluding remarks and future research lines are given in Section 6.

2. Literature Review

The efficiency in education is a mature field of research to the extent that the related literature is plenty of contributions. Some of these studies use parametric methods, such as stochastic frontier analysis (see e.g. [5, 6, 7, 8]) and others consider non-parametric techniques, such as data envelopment analysis DEA (see e.g. [9, 10, 11, 12]). DEA[1] has been broadly accepted and applied to the topic of education (in its distinct variants), representing one of the top five fields of assessment with this methodological tool [14]. A recent contribution [15] provides a wide literature review of different methodological approaches applied to this research field. Additionally, focusing on the performance of students in terms of average test scores in different subjects to face educational efficiency analyses, as we do, we find some recent research works such as those in [16, 17, 18, 19].

An important issue regarding the performance of teachers in education is that related to the concepts of effectiveness[2] and efficiency, as they are closely related. Hence, it is relevant to delimit both concepts. In the literature, many definitions of what an effective teacher is can be found. For example, [21] stated that effective teachers are "(...) those who provide their students with knowledge that is useful in future learning, presumably require their students to exert effort by paying attention and being concentrated in class and by doing demanding homework" (p. 84). [22, 23] highlighted the efficiency dimension of effective teachers, as they claimed that effective teachers are expected "(...) to organize and manage the classroom environment as an efficient learning environment"[3]. Furthermore, they considered that maximizing engagement rates is an expected feature of an effective teacher, who must be able to increase his/her students' interest and implication in their studies; the student engagement refers to the ability of a teacher to reach out to their students using their teaching practices to make them learn [8]. This implies that student engagement plays a relevant role as an outcome of effective teachers.

On the other hand, the DRP methodology applied in this paper has been used in other fields in a satisfying way: to evaluate the sustainability of territories [25], to measure the human development [26], to measure child and maternal health in developing countries [27], to evaluate the competitiveness of the different countries [28], and to stablish a ranking taking into account those aspects related to the ease of doing business in terms of countries' regulation [29]. Precisely, the competitiveness of a country can be

---

[1] [13] paved the grounds for DEA.

[2] [20], among others, reviews the literature about educational effectiveness.

[3] Cited in [24], p. 233.

conceived as the set of institutions, policies and factors that determines the level of productivity of a country. Following this line, to measure the productivity of a country, [30] propose an adaptation of PROMETHEE in which no normalization is considered. They use a partial compensatory approach by means of preference functions defined in the PROMETHEE family. This is similar, to some extent, to our approach although it has the difficulty in delimiting the preference functions of each variable and its parameters (thresholds of preference and indifference).

However, to the best of our knowledge, it is the first time that synthetic indexes based on the DRP approach are applied to evaluate the efficiency of education, more precisely, to assess teachers' efficiency. In Section 4, in addition to describing this methodology step by step, we will enumerate some key aspects and advantages so that this way of measuring the efficiency of teachers is novel and contains very interesting properties that are worth studying.

3. Data

The information analysed in this paper to study the efficiency of Spanish teachers comes from TIMSS and PIRLS 2011, as this was the only year both studies coincided. The purpose of TIMSS is to measure the learning achievement of students in the areas of mathematics and science at the end of the fourth (9–10 years old) and eighth grades (13–14 years old), while PIRLS is focused on the reading achievement of fourth-grade students. Both TIMSS and PIRLS information were gathered by using a two-stage random sample design [31]. Firstly, schools were selected and, secondly, complete classes were sampled within the selected schools. Hence, each classroom (i.e. each student) is associated with the teacher who teaches the corresponding subject. The TIMSS and PIRLS studies consist of four models of context questionnaires: student, home (only in PIRLS), teacher, and school. Other relevant international assessment studies such as PISA are available as well, but the data provided by the latter did not contain information about teachers' characteristics and practices until the 2015 cycle, and this information is not linkable at a student level, only at a school level [32]. Therefore, TIMSS/PIRLS 2011 are the most adequate and recent available data to study teachers' efficiency in Spain. The Spanish sample contains 4,043 students who participated in both tests and were taught by 174 teachers.

Both PIRLS 2011 (for reading) and TIMSS 2011 (for mathematics and science) include an index called "Students Engaged in Reading/Mathematics/Science Lessons index" to represent students' engagement. According to [33], this index was created by asking students the degree to which they agree a lot, agree a little, disagree a little or disagree a lot, with the following 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"[4]. [31] provided a measure of the validity of this index using the Cronbach alpha reliability coefficient and concluded that it reached an acceptable level for Spain (0.68 for reading, 0.6 for mathematics and 0.64 for science). Based on these findings, the engagement index has been considered as a continuous variable in our study as a proxy of teachers' efficiency.

Next, we describe the procedure employed to select the variables of the teaching-learning process for our analysis. We have classified each available teacher-related variable according to its description of teacher efficiency characteristics, based on the definition provided in Section 1. To determine whether a teacher variable represents efficiency, we have asked ourselves the following question: "Does this variable represent the available resources for teachers in order to develop their lessons, such as materials, classroom characteristics, available time or students' background, which are not easily alterable by the teacher?" The selected variables that answered to this question and had a significant effect on explaining teacher efficiency were included in this study. The set of variables considered are included in Table 1, which also shows their descriptive statistics.

Table 1 Selected variables and their descriptive statistics from TIMSS/PIRLS 2011 in fourth grade, Spain

Variables	Notation	Type	Mean	Std. Dev.	Max	Min
Mean scores in:						
Reading	v_1	Cont.	519.18	33.54	592.29	392.96
Mathematics	v_2	Cont.	489.22	36.52	581.62	347.87
Science	v_3	Cont.	512.81	36.00	604.63	376.26
Engagement in:						
Reading	v_4	Cont.	9.94	0,87	12.16	6.48
Mathematics	v_5	Cont.	10.11	0.85	12.64	5.81
Science	v_6	Cont.	10.08	0.87	12.05	6.10
Mean of home resources for learning index	v_7	Cont.	10.30	1.07	12.96	7.34
Proportion of female students	v_8	Cont.	0.50	0.12	1.00	0.19
Proportion of student with diglossia	v_9	Cont.	0.09	0.15	1.00	0.00
Teaching minutes in:						
Reading	v_{10}	Cont.	337.24	104.22	780.00	120.00
Mathematics	v_{11}	Cont.	278.02	42.79	420.00	180.00
Science	v_{12}	Cont.	237.96	51.53	540.00	60.00

Source: Author's own elaboration.

[4] PIRLS 2011 includes two additional questions in the engagement index: "I like what I read about in school" and "My teacher gives me interesting things to read".

4. Methodology

In a broad sense, under the paradigm of multiple criteria decision making, the final aim when solving a multi-criteria optimization problem is the analysis of the possible alternatives available and, taking into account different criteria simultaneously, the selection of the best one, by classifying or ranking them (our case). The purpose of this paper is to calculate an overall measure of the teachers' efficiency from a multiple criteria decision-making perspective, considering the values of all the significant variables. In our case, the set of available alternatives is the set of teachers under study, while the criteria are precisely the significant variables showed in Table 1. Nonetheless, taking into account information about how far is each teacher from satisfying potential desirable values for the variables would greatly enrich the efficiency score that measures the goodness of each teacher. This would allow us to introduce external preferences into the ranking process, enabling the teachers to be analysed according to the desired expectations for the observed data.

To achieve this and to generate a ranking of the teachers according to their efficiency, considering certain desirable values in the process, we build a set of synthetic indicators based on the DRP approach. As explained hereafter, these synthetic indicators suitably combined the values achieved by the teachers for the variables considered with the preference information incorporated by means of two reference levels for each variable: a reservation level (a value considered as acceptable), and an aspiration level (a value regarded as desirable). Furthermore, the synthetic indicators used in our model allow different compensation degrees among the values achieved in the variables considered. This means that bad performances in some variables can be compensated by good performances in others, if desired, giving us the opportunity to obtain the efficiency ranking of teachers taking into account the desired balance among the variables. Additionally, a non-compensatory scheme is also analysed using these synthetic indicators, driving the attention towards the worst variable value achieved by each teacher, which can be useful if we want to develop improvement policies.

4.1. Multi-criteria Decision Making and the Double Reference Point Scheme

To introduce the methodology of the double reference point scheme, we will first define a problem of multi-objective optimization:

$$\begin{aligned} &\max \ \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\} \\ &\text{subject to} \ \ \mathbf{x} \in X, \end{aligned} \tag{1}$$

where $f_j: \mathbb{R}^n \longrightarrow \mathbb{R}$, for $j = 1, \dots, m$, are the j (with $j \geq 2$) conflicting objective functions to be simultaneously optimized over the feasible set $X \subset \mathbb{R}^n$, which is constituted by feasible decision vectors denoted by $\boldsymbol{x} = (x_1, \dots, x_n)^T$. Their images in the objective space, given by $f(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\dots,f_m(\boldsymbol{x}))^T$ for any $\mathbf{x} \in X$, are referred to as objective vectors and form the so-called feasible objective region $Z = f(X) \subset \mathbb{R}^m$. In case X is constituted by a finite number of solutions, the problem is usually refer to as a *multi-criteria optimization problem* (*discrete case*), while we refer to it as a *multi-objective optimization problem* (*continuous case*) if there is an infinite set of solutions.

The conflict degree among the objectives makes it impossible to find a feasible solution minimizing all of them at the same time, and much attention is given to the so-called Pareto optimal solutions representing different trade-offs among the objectives. At these solutions, none of the objectives can be improved without deteriorating, at least, one of the others. Formally, a decision vector $\mathbf{x}' \in X$ is said to be *efficient* or *Pareto optimal* if there does not exist any other $\mathbf{x} \in X$ such that $f_j(\mathbf{x}') \leq f_j(\mathbf{x})$ for every $j = 1, \dots, m$ and $f_i(\mathbf{x}') < f_i(\mathbf{x})$ for, at least, one index *i*.

Nevertheless, Pareto optimal solutions are mathematically incomparable without the consideration of additional information about the preferences of a decision maker (DM). In this regards, the reference point preferential scheme offers a very easy way to provide the preferences. The DM just needs to specify an *aspiration* or *reference value* q_j that (s)he desires for the objective function f_j $(j = 1, \dots, m)$, which constitutes the component j of a so-called *reference point* $\boldsymbol{q} \in \mathbb{R}^m$, i.e. $\boldsymbol{q} = (q_1, \dots, q_m)^T$. In this approach, an achievement scalarizing function is usually formulated to combine the original multi-objective optimization problem and the preferential information given by means of the reference point. By maximizing the achievement scalarizing function over the feasible set, an efficient solution is obtained, which somehow constitutes the most suitable alternative to optimize all the criteria at the same time according to the preferences of the DM. See [34, 35] for further details.

Although the reference point methodology was firstly suggested in [36], the double reference point (DRP) preferential scheme was introduced twenty years later in [4]. Namely, the DM is asked to give, for each $j = 1, \dots, m$, a *reservation level* q_j^r (i.e. a value below which the value of the objective f_j is considered unacceptable), and an *aspiration level* q_j^a (with the same meaning as the reference values previously described, i.e. desirable values for the objective function f_j). Then, this scheme, which can be applied for both the continuous and discrete cases, considers a linear piecewise achievement scalarizing function that normalizes the objective values (continuous case) or criteria values (discrete case) achieved taking into account the ranges defined by the aspiration and reservation levels. [4] also suggested the use of this function as a means of ranking different alternatives or solutions according to the achievement of their objective values with respect to the aspiration and reservation levels.

4.2. Building Synthetic Indexes for the Analysis of Teachers Efficiency

The methodology used to build the synthetic indexes is described in detail in [25] and, as already mentioned, it is based on the DRP preferential scheme introduced by [4]. Here, we describe the procedure followed to calculate these synthetic indexes for the study of the teachers' efficiency.

Let us denote by N_T the number of teachers considered (alternatives) and by N_V the number of variables used to evaluate them (criteria). As previously mentioned, 174 teachers constitute our sample and the 12 variables described in Table 1 are used, which are related to their students' performance in reading, mathematics and science, their classroom characteristics and the available time for each subject. This means that we are considering a multi-criteria optimization problem with a feasible set formed by N_T = 174 alternatives (the teachers in our sample) whose efficiency is analysed according to $N_V =$ 12 criteria (the variables in Table 1).

For each $i = 1, \dots, N_T$ and $j = 1, \dots, N_V$, let x_{ij} be the value of the teacher i for the variable v_j. For each variable, it is necessary to determine whether it is of the "more is better" type (equivalent to maximizing in the continuous case) or if it belongs to the "less is better" type (equivalent to minimizing in the continuous case). In our case, seven variables are of the "more is better" type (mean scores in reading/mathematics/science, engagement in reading/ mathematics/science, and proportion of female students; that is, $v_1, \dots, v_6$ and v_8), and five variables are of the "less is better" type (mean of the home resources for learning index, proportion of students with diglossia, and teaching minutes in reading/mathematics/science; that is, $v_7, v_9 \dots, v_{12}$). Let us denote by q_j^M and q_j^m, respectively, the maximum and minimum values taken by each variable for all the teachers in the sample, that is:

$$q_j^m = \min_{i=1,\dots,N_T} x_{ij}, \quad \forall j = 1, \dots, N_V, \tag{2}$$

$$q_j^M = \max_{i=1,\dots,N_T} x_{ij}, \quad \forall j = 1, \dots, N_V. \tag{3}$$

To build the synthetic indexes based on the DRP scheme, the following steps are performed:

1. We need to set an aspiration and a reservation level for each variable v_j ($j = 1, \dots, N_V$), denoted by q_j^a and q_j^r, respectively. These preferential levels play an important role for the indexes' building process and for the understanding of the results, as will be shown hereafter. Section 4.3 provides details about how they have been set.

2. For each $i = 1, \dots, N_T, j = 1, \dots, N_V$, the value x_{ij} achieved by the teacher i for the variable v_j is normalized with respect to its reservation and aspiration levels, so that we scale all the teachers' variable values and obtain their "position" regarding the corresponding reservation-aspiration ranges. However, this normalization is performed in a different way for variables of the "more is better" type and of the "less is better" type.

On the one hand, for the "more is better" type variables, it is hold that $q_j^m \leq q_j^r \leq q_j^a \leq q_j^M$ and the normalized value of the variable v_j for the teacher i, referred to as y_{ij}, is calculated depending on the "position" of x_{ij} regarding these values:

- If $q_j^m \leq x_{ij} \leq q_j^r$, then $y_{ij} = \frac{x_{ij}-q_j^r}{q_j^r-q_j^m}$.
- If $q_j^r \leq x_{ij} \leq q_j^a$, then $y_{ij} = \frac{x_{ij}-q_j^r}{q_j^a-q_j^r}$. (4)
- If $q_j^a \leq x_{ij} \leq q_j^M$, then $y_{ij} = 1 + \frac{x_{ij}-q_j^a}{q_j^M-q_j^a}$.

On the other hand, for the "less is better" type variables, we have $q_j^m \leq q_j^a \leq q_j^r \leq q_j^M$ and the normalized values y_{ij} are obtained as follows:

- If $q_j^m \leq x_{ij} \leq q_j^a$, then $y_{ij} = 1 + \frac{x_{ij}-q_j^a}{q_j^m-q_j^a}$.
- If $q_j^a \leq x_{ij} \leq q_j^r$, then $y_{ij} = \frac{x_{ij}-q_j^r}{q_j^a-q_j^r}$. (5)
- If $q_j^r \leq x_{ij} \leq q_j^M$, then $y_{ij} = \frac{x_{ij}-q_j^r}{q_j^r-q_j^M}$.

This piecewise normalization implies that, for any type of variable, the normalized variable value y_{ij} takes a negative value (between -1 and 0) if the value x_{ij} of the variable v_j achieved by the teacher i is worse than its reservation level q_j^r. A value of y_{ij} between 0 and 1 means that x_{ij} performs better than the reservation level q_j^r, but worse than its aspiration level q_j^a. Finally, if y_{ij} is between 1 and 2, then x_{ij} is better than the aspiration level q_j^a.

3. Now, we can build a *weak index* WI_i for each teacher i, which is the arithmetic mean of the normalized variable values:

$$\mathrm{WI}_i = \frac{1}{N_V}\sum_{j=1}^{N_V} y_{ij}, \quad \forall i = 1, \dots, N_T. \tag{6}$$

Note that WI_i also takes values between −1 and 2, and therefore, it can be understood as the "position" of the teacher with respect to hypothetical global reservation and aspiration levels.

4. In addition, a *strong index* SI_i can be obtained to measure the worst performance of each teacher i. It is defined as the minimum of all the normalized variable values:

$$\mathrm{SI}_i = \min_{j=1,\dots,N_V} y_{ij}, \ \forall i = 1, \dots, N_T. \tag{7}$$

A negative value of SI_i for a teacher i means that, at least, there is one variable value of this teacher that performs worse than its corresponding reservation level (i.e. at least, there exists one normalized variable value below 0). Furthermore, a value of SI_i over 1 means that, for the teacher i, all variables have reached better values than their respective aspiration levels (i.e. all normalized variable values are above 1).

It is important to remark that, while the weak index allows full compensation among all the variables (substitutability, i.e. a poor performance in one variable can be compensated by a good performance in another one), the strong index does not enable any compensation among them since it represents the worst value achieved by the normalized variables values. Observe that $\mathrm{SI}_i \leq \mathrm{WI}_i, \forall i = 1, \dots, N_T$.

5. In view of the fact that the weak and strong indexes represent extreme situations, enabling either full compensation or no compensation, a combination of them both can be used to enable a partial compensation among the values achieved by the variables. Thus, a *mixed synthetic index* MI_i can be obtained for each teacher i as a linear combination of them:

$$\mathrm{MI}_i = \mu\, \mathrm{WI}_i \ + (1-\mu)\, \mathrm{SI}_i, \quad \forall i = 1, \dots, N_T, \tag{8}$$

with $0 \leq \mu \leq 1$ being a real value. Note that $\mathrm{MI}_i = \mathrm{SI}_i$ in case $\mu = 0$ (no compensation), while $\mathrm{MI}_i = \mathrm{WI}_i$ if $\mu = 1$ (full compensation). Any value of μ between 0 and 1 reflects an intermediate state between these two extremes, taking into account that the larger the value of μ, the more compensation is allowed. Note that $\mathrm{SI}_i \leq \mathrm{MI}_i \leq \mathrm{WI}_i, \forall i = 1, \dots, N_T$.

Traditionally, weights have been used in the reference point schemes with a normalization role, but it is also possible to use weights with a preferential meaning [35]. Note that WI_i, SI_i and MI_i can incorporate a weight factor for each variable, if desired. Provided that different weights want to be assigned to the variables, the building process would be adapted as follows. For each $j = 1, \dots, N_V$, let us denote by ω_j the weight value given to the variable v_j ($\omega_j > 0$), and let us consider their normalized values as $\bar{\omega}_j = \frac{\omega_j}{\sum_{k=1}^{N_V} \omega_k}$ (observe that the normalized weights add up to 1). The *weighted weak index* W-WI_i of teacher i would be calculated as the weighted sum:

$$\text{W-WI}_i = \sum_{j=1}^{N_V} \bar{\omega}_j\, y_{ij}, \quad \forall i = 1, \dots, N_T. \tag{9}$$

For building the strong index and avoiding unwanted effects due to the use of weights and to the fact that y_{ij} can take positive and negative values, additional mathematical transformations are required. Let us consider the following modified weights:

$$\hat{\omega}_j = \frac{\bar{\omega}_j}{\max\limits_{k=1,\dots,N_V} \bar{\omega}_k}, \ \forall j = 1, \dots, N_V. \tag{10}$$

For each teacher i and each variable v_j, we transform the values y_{ij} in the following way:

$$\bar{y}_{ij} = y_{ij} - m_i\,, \quad \text{where} \quad m_i = \left[\min_{k=1,\dots,N_V} y_{ik}\right] + 1. \qquad (11)$$

The $[\,\cdot\,]$ operator gives the integer part of any real number. Now, the *weighted strong index* W-SI$_i$ of teacher i would be redefined as:

$$\text{W-SI}_i = m_i + \min_{j=1,\dots,N_V} \widehat{\omega}_j \bar{y}_{ij}\,, \quad \forall i = 1, \dots, N_T. \qquad (12)$$

Obviously, the mixed index would be updated accordingly using the weighted weak and strong indexes. That is, the *weighted mixed index* W-MI$_i$ of teacher i would be:

$$\text{W-MI}_i = \mu\, \text{W-WI}_i + (1-\mu)\, \text{W-SI}_i, \quad \forall i = 1, \dots, N_T. \qquad (13)$$

where μ is the compensation grade considered.

As just explained, the DRP approach can be used to produce compensatory or non-compensatory synthetic indexes to study the teachers' efficiency, or even mixed indexes for a compensation grade. Let us to point out some properties and key aspects of this kind of normalization-aggregation methodology:

1) The DRP normalization applied is more general than the commonly used range normalization (consisting on dividing each variable value by the difference among the maximum and minimum values achieved), in which 0 means the worst value attained and 1 is associated with the best one. Indeed, if we consider the minimum and maximum values as the aspiration and the reservation levels in the DRP scheme, the range normalization would be a particular case of the DRP normalization.

2) This normalization allows us to include preferences into the process, which regards to potential desirable ranges for the variable values (i.e. the aspiration and the reservation levels), enabling the efficiency of each teacher to be evaluated also depending on his/her performance regarding these values.

3) Each unit of the normalized value represents the magnitude in each sub-interval defined by the aspiration and the reservation levels, and the minimum and maximum values attained by each variable. Actually, the value of each normalized variable has an understandable meaning in practice, since it indicates the position (i.e. the performance of each teacher) with respect to the aspiration and reservation levels (i.e. with respect to the desirable threshold for each variable). Therefore, the DRP-based indexes give us measurements about how far is each teacher from satisfying the potential desirable values for the variables (using different compensation schemes).

4) The indexes built make it possible to set weights representing the relative importance of each individual variable, if desired, allowing us to give more or less importance to each input and output.

5) To measure the efficiency of the teachers, three ways of aggregation (the three indexes) are available, taking into account different compensation schemes between the variables (outputs and inputs). One weak index, which enables absolute compensation among variables. A strong index that does not allow any compensation among the

variables and that measures the worst (normalized) variable value. Finally, a compensation between them both is possible through the mixed index, calculated as a linear combination of both indexes (for a certain compensation grade).

6) These indexes enable us to build an efficiency ranking of the teachers under a wider multi-criteria optimization perspective, rather than following the conventional scheme in which just the outputs are maximized, or just the inputs are minimized (for specific input or output values, respectively). The efficiency is measured assuming that all the variables (inputs and outputs) must be simultaneously optimized, in order to determine which policies should be promoted in the future by politicians or stakeholders to reach a teaching environment enabling the teachers to be efficient. Under this perspective, we can have a better insight about how to distribute the available resources to improve key factors of the teaching-learning context that are desirable to let teachers achieve efficiency (as an optimal balance between inputs and outputs).

4.3. Aspiration and Reservation Values Used for Assessing the Teachers' Efficiency

For the normalization of the variable values carried out in (4) and (5) and for the building process, key parameters are the aspiration and reservation levels, as above mentioned. Let us recall that they represent, respectively, a desirable level to be achieved by each variable and a level below which the variable values are regarded as unwanted. These preferential levels can be set in several ways. For example, they may be defined using absolute values, but such universally accepted values do not exist in the literature for the variables under study in this paper. In the same line, another option may be to involve a panel of experts to give appropriate values according to their experience and expectations, but some inconsistencies might arise. In addition, these values may be set in a relative manner, i.e. taking into account the situation of one or several teachers with respect to the others for each variable.

However, in this study, the aspiration and reservation levels have been set based on the statistical descriptors shown in Table 1, in order to use realistic values according to the observational data considered. To be more precise, for each of the "more is better" type variables, we have considered the mean value plus the standard deviation as the aspiration value, and the mean value minus the standard deviation as the reservation value. Likewise, for the "less is better" type variables, the aspiration values have been set to their mean minus the standard deviation, while the reservation values are their mean plus the standard deviation. Table 2 contains the aspiration and reservation levels used.

5. Results and Discussion

In this section, we analyse the efficiency of the Spanish teachers according to the ranking obtained from the synthetic indexes described in Section 4.

Table 2 Aspiration and reservation levels of the selected variables

Variables	Notation	"More is better" or "Less is better" type	Aspiration level	Reservation level
Mean scores in:				
Reading	v_1	"More"	553.42	486.33
Mathematics	v_2	"More"	525.74	452.70
Science	v_3	"More"	548.81	476.81
Engagement in:				
Reading	v_4	"More"	10.80	9.07
Mathematics	v_5	"More"	10.96	9.26
Science	v_6	"More"	10.95	9.21
Mean of home resources for learning index	v_7	"Less"	9.23	11.36
Proportion of female students	v_8	"More"	0.62	0.37
Proportion of student with diglossia	v_9	"Less"	0.00*	0.29*
Teaching minutes in:				
Reading	v_{10}	" Less"	233.02	441.45
Mathematics	v_{11}	" Less"	235.22	320.81
Science	v_{12}	" Less"	186.43	289.49

Source: Author's own elaboration.

* Altered values.

5.1. Calculation of the Weak, Strong and Mixed Indexes

For the determination of the weak, strong and mixed indexes, we have used the same weights for all the variables in order to give the same importance to all of them. Thus, the weak and strong indexes have been obtained according to expressions (6) and (7), respectively. For the mixed index, we have considered $\mu = 0.5$ as the compensation grade in expression (8).

For the sake of clarity, Tables 3 and 4 report the results for the Spanish teachers in our sample, classified in the first (Q1) and the fourth quartile (Q4), respectively, once they have been ranked in descending order according to the mixed synthetic index[5]. We have decided to study the teachers ranked according to the mixed index since it represents a combination of the weak and strong indexes, and these two represent the extreme degrees of compensation. These tables provide an overall view, enabling us to have a broad understanding of the most efficient teachers and the least ones, and to identify key variables. To this end, the mean and the standard deviation of the values achieved by the three indexes and the normalized variable values are given in the first two rows.

[5] The weak, strong and mixed indexes for the rest of Spanish teachers are available upon request.

In addition to the values reached for the indexes, the ranking position of each teacher –regarding each index– is also indicated in the first three columns (in bold face to highlight the order considered to rank the teachers). As said, Tables 3 and 4 also depict the normalized variable values attained by the teachers shown, which are calculated following expressions (4) and (5). Normalized variable values above 1 are colored in light grey, while these between 0 and 1 are colored in medium grey, and lastly the ones lower than 0 are colored in dark grey. In this way, for each variable, we can easily visualize at a glance teachers performing better than the aspiration level (in light grey), between the reservation and aspiration levels (in medium grey), or worse than the reservation level (in dark grey). Finally, the worst variable value of each teacher (which corresponds to the strong index) is highlighted with a bold-faced box.

According to Table 3, in general, the teachers at the highest positions fulfill an acceptable performance for all the variables. These teachers are mainly centered around the values 0.439, 0.158 and 0.739, for the mixed, strong and week index values, respectively; observe that $0.158 \leq 0.439 \leq 0.719$. The low but positive mean value obtained for the strong index implies that the most efficient teachers (regarding the mixed index) got low values for the worst variable, but above 0. Then, overall, their worst variable is still among the desirable limits. Note that the variables regarding the scores and the engagement levels reached by the students in the three subjects, and the proportion of students with diglossia, achieve the highest mean variable values. This indicates its important role for the efficiency of the teachers.

The most efficient teacher (according to the mixed index) has five variable values above their aspiration levels, while the other ones are within the aspiration and the reservations levels. Thus, it seems that the outstanding performance of this teacher regarding the mean scores and engagement levels in mathematics and science of his/her students, and the proportion of female students in his/her classrooms were crucial for being ranked as the most efficient teacher. However, being between the desirable limit for the rest of variables was also very relevant[6]. Nevertheless, the most interesting finding is that his/her worst variable behavior (reached for the teaching minutes in science) is the best one in comparison to the rest of teachers.

The teacher in the second position in Table 3 gets a mixed index that significantly differs from the first teacher, although (s)he does not show a significant variation with the next teachers, which can be understood as if they have a similar efficiency level. Observe that the ranking of these teachers according to the weak and strong indexes fluctuates somehow, which implies that their full balanced and worst performances show some variability.

Regarding the teachers in Q4, as expected, a high number of dark grey cells can be seen in Table 4, indicating a poor performance (i.e. under their reservation levels) of

[6] Note that this teacher also got the first position for the weak and strong indexes, which implies that (s)he is also the most efficient for a full compensation and for no compensation among the variables, respectively.

some variables. This is also reflected by the negative mean values reached by two of the synthetic indexes (-0.251 for the strong and -0.148 for the mixed), although the weak index got a positive mean value (0.356, due to the full compensation among the normalized variable values). Moreover, although the means of the variable values are between the desirable ranges defined by the aspiration and reservation levels, they tend to be below the mid value among the desirable limits.

In Table 4, we can observe the teachers in positions 142, 148, 150; they perform over the aspiration levels for the scores –in the three subjects– and within the desirable ranges of values for almost all the rest of variables. Only one or two variables showed a bad performance below the reservation levels, among which we always find the index measuring the home resources for learning. This explains that these teachers are at the bottom of the efficiency ranking, despite of their acceptable performance regarding the majority of variables. Therefore, a low level of learning resources at the students' home has a negative impact for a teacher to be efficient, and this seems to be difficult to be compensated with good performances in other dimensions.

The least efficient teacher, ranked in the last position in Table 4, attains the worst possible mean score value in the three subjects, thus making its strong index to reach the lowest possible value (–1).

Additionally, the summary descriptive statistics reported in Table 5 (using the original variable values) and in Table 6 (using the normalized variable values), for the top and bottom deciles and quartiles of the teachers distribution ranked according to their mixed index, allow us to withdraw some further intuition regarding our results. The mean scores and engagement indexes reached by the students of those teachers ranked within the top decile/quartile –in the three subjects–, as well as the proportion of female students, are clearly above the figures found for those teachers at the bottom of the efficiency distribution. Conversely, the less efficient teachers deal with students coming from homes with low levels of resources for learning and they are highly "affected" by diglossia in their classrooms, which indicates the importance of these variables for the teachers' efficiency ranking. What is more, this gap between highly efficient and low efficient teachers reaches a standard deviation in terms of the scores, what translate to around one year of the teaching-learning process according to PISA. Likewise, the teaching times for the teachers at the bottom of the efficiency distribution is significantly higher than these for the most efficient ones, particularly for the reading competence, which has important implications in terms of educational policies.

Table 3 Weak, Strong and Mixed indexes and normalized variables values for the Spanish teachers in the first quartile (Q1).

Ranking			**Index**			**Mean scores**			**Engagement**			**Mean of Home Resources for Learning Index**	**Proportion of Female Students**	**Proportion of Students with Diglossia**	**Teaching Minutes**		
Weak	**Strong**	**Mixed**	**Weak**	**Strong**	**Mixed**	**Reading**	**Maths**	**Sciences**	**Reading**	**Maths**	**Sciences**				**Reading**	**Maths**	**Sciences**
Mean			**0.719**	**0.158**	**0.439**	**0.754**	**0.714**	**0.727**	**0.835**	**0.824**	**0.814**	**0.543**	**0.591**	**0.862**	**0.681**	**0.671**	**0.615**
Std. Dev.			**0.121**	**0.131**	**0.081**	**0.361**	**0.415**	**0.394**	**0.358**	**0.277**	**0.371**	**0.390**	**0.426**	**0.194**	**0.336**	**0.429**	**0.286**
1	1	**1**	1.026	0.480	0.753	0.898	1.091	1.417	0.833	1.169	1.110	0.831	2.000	1.000	0.535	0.944	0.480
4	8	**2**	0.959	0.243	0.601	1.614	1.455	1.460	1.546	1.028	0.913	0.322	0.775	1.000	0.679	0.243	0.480
17	2	**3**	0.765	0.417	0.591	0.884	0.964	1.048	0.810	0.598	0.750	0.612	0.417	1.000	0.679	0.944	0.480
9	4	**4**	0.822	0.312	0.567	0.750	1.186	1.247	1.334	0.865	0.910	0.469	0.312	0.828	0.535	0.944	0.480
6	8	**5**	0.867	0.243	0.555	1.385	1.715	1.708	0.481	0.822	0.438	0.258	1.120	1.000	0.751	0.243	0.480
25	3	**6**	0.715	0.381	0.548	0.381	0.456	0.501	0.779	0.524	0.484	0.588	0.695	0.843	1.336	0.944	1.051
10	7	**7**	0.815	0.244	0.530	1.022	0.755	0.946	1.203	0.664	0.724	0.244	1.120	0.836	0.607	1.185	0.480
5	25	**8**	0.890	0.141	0.515	1.377	1.658	1.404	1.282	0.975	0.975	0.491	0.618	1.000	0.223	0.535	0.141
35	8	**9**	0.688	0.243	0.466	0.403	0.278	0.407	1.197	0.993	0.454	1.077	0.914	0.770	0.895	0.243	0.626
18	21	**10**	0.752	0.178	0.465	0.939	0.982	0.718	0.689	0.869	0.816	0.178	0.272	1.000	0.751	1.185	0.626
12	31	**11**	0.803	0.112	0.457	0.886	0.995	1.239	1.059	1.014	1.110	0.391	0.112	1.000	0.607	0.594	0.626
46	6	**12**	0.652	0.262	0.457	0.466	0.473	0.468	0.751	0.989	1.198	0.726	0.262	1.000	0.439	0.477	0.577
39	8	**13**	0.669	0.243	0.456	0.320	0.429	0.374	0.807	1.282	1.417	0.843	0.433	0.724	0.679	0.243	0.480
36	17	**14**	0.683	0.227	0.455	0.492	0.321	0.383	0.227	0.437	0.320	1.284	1.076	0.828	0.966	0.944	0.917
42	8	**15**	0.663	0.243	0.453	0.698	0.755	0.650	0.843	0.744	0.618	0.451	0.800	1.000	0.679	0.243	0.480
28	20	**16**	0.710	0.179	0.444	0.710	0.549	0.671	0.844	0.962	0.855	0.642	0.179	0.713	0.966	0.944	0.480
50	8	**17**	0.644	0.243	0.444	0.250	0.557	0.388	1.048	0.787	1.083	0.902	0.312	1.000	0.679	0.243	0.480
59	8	**18**	0.628	0.243	0.436	0.568	0.265	0.414	0.578	0.851	0.669	0.776	0.867	1.000	0.535	0.243	0.771
3	90	**19**	0.987	-0.120	0.433	0.919	0.978	0.858	0.804	0.915	1.210	-0.120	0.226	1.000	2.000	2.000	1.051
2	95	**20**	0.997	-0.133	0.432	1.633	1.813	1.729	1.361	1.201	0.969	-0.133	-0.046	1.000	0.679	0.477	1.288
82	5	**21**	0.563	0.297	0.430	0.624	0.297	0.326	0.546	0.758	0.454	0.682	0.312	0.655	0.679	0.944	0.480
26	25	**22**	0.712	0.141	0.426	0.837	0.935	0.864	0.951	0.834	0.650	0.776	0.800	1.000	0.223	0.535	0.141

Source: Author’s own elaboration.

Table 3 (Continued).

Ranking			Index			Mean scores			Engagement			Mean of Home Resources for Learning Index	Proportion of Female Students	Proportion of Students with Diglossia	Teaching Minutes		
Weak	Strong	Mixed	Weak	Strong	Mixed	Reading	Maths	Sciences	Reading	Maths	Sciences				Reading	Maths	Sciences
68	16	**23**	0.597	0.238	0.418	0.662	0.543	0.458	0.396	0.584	0.514	0.238	0.695	0.687	0.679	0.944	0.771
16	45	**24**	0.771	0.055	0.413	0.760	1.188	0.974	0.882	0.869	0.469	0.055	0.949	1.000	0.679	0.944	0.480
73	18	**25**	0.583	0.226	0.405	0.482	0.501	0.562	0.764	0.639	0.435	0.944	0.226	0.836	0.535	0.594	0.480
11	55	**26**	0.811	-0.005	0.403	0.611	0.429	0.413	1.139	0.906	0.883	-0.005	1.076	0.483	1.292	1.638	0.868
47	22	**27**	0.650	0.155	0.402	1.358	0.720	0.994	0.691	0.466	0.929	0.155	0.513	1.000	0.247	0.243	0.480
38	28	**28**	0.676	0.125	0.401	0.964	0.528	0.562	0.125	0.601	0.510	0.775	1.340	1.000	0.343	0.594	0.771
92	8	**29**	0.537	0.243	0.390	0.908	0.783	0.709	0.339	0.383	0.407	0.477	0.513	0.808	0.391	0.243	0.480
8	75	**30**	0.841	-0.064	0.389	0.793	0.964	0.183	1.115	1.526	2.000	-0.064	0.593	0.862	0.823	0.243	1.051
70	19	**31**	0.593	0.184	0.388	0.238	0.184	0.338	0.835	0.535	0.530	1.176	0.395	1.000	0.563	0.886	0.432
44	31	**32**	0.657	0.112	0.385	0.689	0.193	0.655	0.657	0.296	0.708	0.931	0.112	1.000	0.535	0.944	1.169
31	42	**33**	0.700	0.068	0.384	0.482	0.212	0.331	1.428	1.105	0.940	1.548	0.513	0.617	0.679	0.068	0.480
32	44	**34**	0.698	0.067	0.383	0.812	0.764	0.796	0.766	0.424	0.527	0.839	0.067	1.000	0.391	0.944	1.051
15	57	**35**	0.774	-0.025	0.375	0.607	0.627	0.461	1.712	1.453	1.468	0.282	1.166	0.819	-0.025	0.243	0.480
54	31	**36**	0.635	0.112	0.373	0.402	0.720	0.842	0.609	0.986	0.856	0.508	0.112	0.828	0.463	0.243	1.051
52	35	**37**	0.643	0.100	0.372	0.239	0.271	0.367	0.484	1.015	1.426	0.341	1.081	0.100	0.966	0.944	0.480
58	31	**38**	0.630	0.112	0.371	0.463	0.316	0.411	0.582	0.496	0.613	0.960	0.112	0.540	1.071	0.944	1.051
41	42	**39**	0.664	0.068	0.366	0.722	0.556	0.671	1.344	0.980	1.246	1.078	0.312	0.483	0.319	0.068	0.189
23	56	**40**	0.736	-0.020	0.358	1.178	0.987	0.804	1.008	1.080	-0.020	0.141	0.513	1.000	0.751	0.769	0.626
80	24	**41**	0.567	0.148	0.357	0.395	0.397	0.257	0.503	0.621	0.528	0.201	0.148	0.687	1.115	1.276	0.674
71	30	**42**	0.591	0.120	0.355	0.437	0.120	0.211	0.546	0.616	1.209	0.575	0.272	1.000	0.679	0.944	0.480
86	23	**43**	0.550	0.152	0.351	0.490	0.625	0.837	0.326	0.546	0.711	0.152	0.513	1.000	0.679	0.243	0.480
22	60	**44**	0.738	-0.042	0.348	1.434	0.892	0.920	0.534	0.855	0.795	0.317	1.223	1.000	0.679	0.243	-0.042

Source: Author's own elaboration.

Table 4 Weak, Strong and Mixed indexes and normalized variables values for the Spanish teachers in the fourth quartile (Q4).

Ranking			**Index**			**Mean scores**			**Engagement**			**Mean of Home Resources for Learning Index**	**Proportion of Female Students**	**Proportion of Students with Diglossia**	**Teaching Minutes**		
Weak	**Strong**	**Mixed**	**Weak**	**Strong**	**Mixed**	**Reading**	**Maths**	**Sciences**	**Reading**	**Maths**	**Sciences**				**Reading**	**Maths**	**Sciences**
Mean			**0.356**	**-0.651**	**-0.148**	**0.425**	**0.401**	**0.386**	**0.225**	**0.242**	**0.270**	**0.466**	**0.272**	**0.714**	**0.251**	**0.251**	**0.367**
Std. Dev.			**0.152**	**0.259**	**0.129**	**0.683**	**0.596**	**0.610**	**0.493**	**0.422**	**0.471**	**0.768**	**0.626**	**0.452**	**0.523**	**0.577**	**0.442**
140	122	**131**	0.352	-0.342	0.005	-0.312	-0.144	-0.342	0.657	0.902	0.727	0.968	0.112	-0.155	0.391	0.944	0.480
154	117	**132**	0.294	-0.286	0.004	-0.286	0.518	0.255	0.204	0.378	0.161	0.074	0.035	0.507	0.391	0.243	1.051
130	127	**133**	0.403	-0.395	0.004	0.443	0.552	0.527	0.402	0.364	0.411	0.926	-0.081	0.754	-0.114	-0.395	1.051
143	121	**134**	0.344	-0.337	0.003	-0.034	-0.131	-0.161	0.786	0.307	0.521	0.996	1.068	-0.337	0.391	0.243	0.480
131	127	**135**	0.399	-0.395	0.002	0.742	0.920	0.747	0.483	0.857	0.824	0.370	-0.081	0.261	0.103	-0.395	-0.042
153	120	**136**	0.301	-0.310	-0.005	0.817	0.302	0.498	-0.310	-0.076	-0.071	-0.220	0.914	1.000	-0.025	0.594	0.189
126	143	**137**	0.412	-0.459	-0.023	1.312	1.133	1.134	-0.227	-0.313	-0.459	-0.038	0.426	0.700	0.751	-0.244	0.771
123	149	**138**	0.438	-0.489	-0.026	0.568	0.445	0.688	0.330	0.651	0.792	-0.489	0.513	1.000	-0.025	0.594	0.189
110	151	**139**	0.476	-0.561	-0.042	1.147	0.676	0.928	-0.561	-0.273	0.037	0.077	0.433	0.862	0.966	0.944	0.480
147	139	**140**	0.331	-0.420	-0.044	-0.420	-0.302	-0.216	0.645	-0.051	0.301	2.000	-0.210	1.000	0.679	0.068	0.480
146	140	**141**	0.333	-0.423	-0.045	-0.244	0.215	0.051	0.762	0.904	0.648	0.944	-0.423	0.594	-0.114	-0.395	1.051
83	154	**142**	0.561	-0.661	-0.050	1.447	1.047	1.205	0.452	0.390	0.744	-0.661	0.800	0.877	-0.291	0.243	0.480
152	138	**143**	0.316	-0.416	-0.050	1.293	0.949	0.746	-0.258	-0.220	-0.416	0.175	0.050	1.000	0.391	-0.395	0.480
155	127	**144**	0.294	-0.395	-0.050	-0.345	-0.320	-0.355	0.856	0.850	0.480	1.580	0.513	1.000	-0.291	-0.395	-0.042
157	125	**145**	0.287	-0.391	-0.052	0.306	-0.229	0.061	0.292	0.392	0.611	0.815	-0.391	1.000	0.391	0.243	-0.042
159	127	**146**	0.283	-0.395	-0.056	-0.186	-0.205	-0.216	0.419	0.067	-0.061	1.290	1.534	0.797	0.391	-0.395	-0.042
161	127	**147**	0.277	-0.395	-0.059	0.969	0.729	0.592	-0.213	-0.136	-0.286	-0.096	0.290	1.000	0.391	-0.395	0.480
20	162	**148**	0.751	-0.874	-0.061	1.356	1.208	1.295	0.920	0.792	1.144	-0.874	0.775	1.000	0.679	0.243	0.480
144	148	**149**	0.341	-0.485	-0.072	0.436	0.516	0.337	-0.321	-0.143	-0.485	0.638	1.076	0.828	0.535	-0.093	0.771
109	153	**150**	0.481	-0.646	-0.082	1.543	1.258	1.276	0.137	0.188	0.077	-0.522	0.035	1.000	-0.646	0.944	0.480
45	161	**151**	0.656	-0.830	-0.087	1.100	0.876	1.076	0.786	0.896	1.071	-0.830	-0.066	0.862	0.679	0.944	0.480
156	144	**152**	0.293	-0.468	-0.088	0.458	-0.243	-0.023	0.048	-0.176	0.599	0.915	-0.468	1.000	0.679	0.243	0.480

Source: Author's own elaboration.

Table 4 (Continued).

Ranking			Index			Mean scores			Engagement			Mean of Home Resources for Learning Index	Proportion of Female Students	Proportion of Students with Diglossia	Teaching Minutes		
Weak	Strong	Mixed	Weak	Strong	Mixed	Reading	Maths	Sciences	Reading	Maths	Sciences				Reading	Maths	Sciences
168	125	**153**	0.207	-0.391	-0.092	-0.189	0.168	-0.355	-0.065	0.332	0.035	0.603	-0.391	0.655	0.966	0.243	0.480
169	127	**154**	0.201	-0.395	-0.097	0.046	-0.051	-0.098	-0.142	0.089	-0.002	1.162	0.407	0.456	-0.114	-0.395	1.051
167	150	**155**	0.223	-0.507	-0.142	-0.312	-0.507	-0.460	0.576	0.638	0.667	1.192	0.112	0.540	-0.202	0.243	0.189
172	142	**156**	0.133	-0.424	-0.145	-0.424	0.427	0.167	-0.350	-0.058	-0.068	0.510	-0.104	-0.089	0.679	0.944	-0.042
51	164	**157**	0.644	-1.000	-0.178	0.479	1.463	0.768	0.865	0.184	0.921	0.623	1.340	-1.000	0.751	0.418	0.917
61	164	**158**	0.626	-1.000	-0.187	0.862	0.677	0.860	1.180	0.948	0.720	0.310	1.027	1.000	0.679	0.243	-1.000
66	164	**159**	0.606	-1.000	-0.197	2.000	2.000	2.000	-0.135	0.008	-0.114	-0.588	-1.000	1.000	0.679	0.944	0.480
132	160	**160**	0.396	-0.823	-0.214	0.216	0.545	0.503	0.180	0.212	0.135	0.957	0.914	0.483	-0.823	0.944	0.480
166	155	**161**	0.228	-0.662	-0.217	0.074	0.284	-0.278	0.349	-0.141	0.225	1.244	-0.662	0.138	0.823	-0.093	0.771
81	164	**162**	0.564	-1.000	-0.218	0.657	0.710	0.714	1.009	0.740	1.155	0.863	1.068	1.000	0.247	-1.000	-0.401
150	158	**163**	0.329	-0.812	-0.242	-0.154	-0.017	0.123	0.110	-0.057	0.306	1.631	-0.812	1.000	0.391	0.944	0.480
171	156	**164**	0.160	-0.682	-0.261	-0.682	-0.647	-0.666	0.134	0.013	-0.152	1.082	0.513	1.000	0.607	0.243	0.480
125	163	**165**	0.421	-0.966	-0.272	0.891	0.698	0.946	0.146	0.609	-0.064	-0.966	0.667	0.867	0.535	0.243	0.480
121	164	**166**	0.455	-1.000	-0.273	1.594	0.830	0.973	0.144	0.343	0.399	-1.000	-0.932	1.000	0.679	0.944	0.480
165	159	**167**	0.243	-0.816	-0.286	-0.077	0.623	0.422	-0.816	0.276	0.122	-0.257	-0.662	1.000	0.823	1.276	0.189
145	164	**168**	0.336	-1.000	-0.332	0.671	0.793	0.641	-0.008	0.363	0.361	-0.144	0.433	1.000	0.679	0.243	-1.000
149	164	**169**	0.329	-1.000	-0.335	0.418	-0.004	0.149	0.868	0.542	0.459	0.804	0.513	1.000	-1.000	0.243	-0.042
151	164	**170**	0.326	-1.000	-0.337	0.350	-0.101	0.044	0.389	0.406	0.725	0.983	0.914	1.000	-1.000	0.243	-0.042
174	157	**171**	0.072	-0.793	-0.361	-0.276	-0.135	-0.219	-0.301	-0.208	-0.232	1.086	-0.104	-0.171	-0.793	1.638	0.577
162	164	**172**	0.264	-1.000	-0.368	0.698	0.670	0.711	-0.151	-0.372	-0.152	0.059	1.120	1.000	0.103	-1.000	0.480
170	164	**173**	0.163	-1.000	-0.418	0.761	0.431	0.923	-1.000	-1.000	-1.000	0.700	0.513	1.000	0.247	-0.093	0.480
173	164	**174**	0.108	-1.000	-0.446	-1.000	-1.000	-1.000	0.615	0.243	0.058	1.596	0.262	1.000	-0.202	0.243	0.480

Source: Author's own elaboration.

Table 5 Descriptive statistics of teachers in the first decile (D1), first quartile (Q1), tenth decile (D10), and the fourth quartile (Q4), ordered by the mixed index, using the original variable values.

		Mean scores			Engagement			Mean of Home Resources for Learning Index	Proportion of Female Students	Proportion of Students with Diglossia	Teaching Minutes		
		Reading	Maths	Sciences	Reading	Maths	Sciences				Reading	Maths	Sciences
D1	Mean	537.19	513.39	538.06	10.64	10.73	10.63	10.08	0.55	0.02	296.18	265.59	232.65
	Stand. Desv.	21.92	28.65	28.52	0.50	0.37	0.47	0.62	0.14	0.03	44.87	29.35	19.86
Q1	Mean	534.18	503.33	527.59	10.50	10.66	10.57	10.20	0.52	0.04	303.58	264.19	224.07
	Stand. Desv.	20.07	28.00	26.21	0.57	0.47	0.56	0.82	0.12	0.06	60.18	33.25	28.74
D10	Mean	508.52	479.68	502.94	9.30	9.38	9.45	10.23	0.46	0.10	429.44	292.50	285.28
	Stand. Desv.	48.59	53.47	53.02	1.17	1.10	1.09	1.51	0.17	0.24	180.30	58.03	99.09
Q4	Mean	509.48	478.22	501.00	9.36	9.54	9.56	10.31	0.45	0.10	405.80	301.70	260.57
	Stand. Desv.	46.21	46.72	47.01	1.00	0.95	1.01	1.48	0.15	0.19	140.88	50.73	72.60

Source: Author's own elaboration.

Table 6 Descriptive statistics of teachers in the first decile (D1), first quartile (Q1), tenth decile (D10), and the fourth quartile (Q4), ordered by the mixed index, using the normalized variable values.

		Mean scores			Engagement			Mean of Home Resources for Learning Index	Proportion of Female Students	Proportion of Students with Diglossia	Teaching Minutes		
		Reading	Maths	Sciences	Reading	Maths	Sciences				Reading	Maths	Sciences
D1	Mean	0.79	0.86	0.88	0.93	0.87	0.83	0.61	0.67	0.91	0.71	0.65	0.55
	Stand. Desv.	0.39	0.44	0.44	0.32	0.22	0.29	0.30	0.47	0.11	0.24	0.35	0.19
Q1	Mean	0.74	0.71	0.72	0.84	0.82	0.81	0.55	0.58	0.86	0.68	0.68	0.63
	Stand. Desv.	0.35	0.42	0.40	0.36	0.28	0.37	0.39	0.42	0.19	0.34	0.43	0.27
D10	Mean	0.42	0.43	0.42	0.20	0.17	0.22	0.50	0.28	0.74	0.19	0.37	0.24
	Stand. Desv.	0.71	0.68	0.68	0.57	0.44	0.48	0.79	0.76	0.54	0.64	0.67	0.53
Q4	Mean	0.43	0.40	0.39	0.22	0.24	0.27	0.47	0.27	0.71	0.25	0.25	0.37
	Stand. Desv.	0.68	0.60	0.61	0.49	0.42	0.47	0.77	0.63	0.45	0.52	0.58	0.44

Source: Author's own elaboration.

6. Conclusions

In this paper, we have proposed to study the efficiency of primary education teachers according to a set of synthetic indexes built using multi-criteria optimization techniques, in concrete, the double reference point approach. To this aim, data from TIMSS and PIRLS 2011 for fourth-grade reading, mathematics and science teachers in Spain have been used. The synthetic indexes calculated gathering these data are: a weak index allowing full compensation among the variables, a strong index informing about the worst variable value (no compensation), and a mixed synthetic index representing a different degree of compensation of the two formers. As a novelty, these indexes are built using aspiration and reservation levels in the normalization of each variable (output and input), which define desirable ranges of values to be achieved by the variables used to study the teachers' efficiency.

In particular, low efficiency teachers are related to a high number of teaching hours. In other words, the teaching times for the teachers at the bottom of the efficiency distribution are significantly higher than the times employed for teaching by the most efficient ones. This is in line with some recent literature that showed that, in general, weekly instruction time has no effect on children's academic achievement for Spain [37][7]. These results may indicate that the frequently used procedure in Spain of having children exposed to a large amount of weekly instruction time may not be useful to increase their academic achievement. Furthermore, this excessive amount of instruction time may cause an increase in the monetary costs of education [38] –teachers' salaries, school resources, etc.– and time costs –opportunity cost of this time, limitation in the selection of activities that children can do during the day, etc.–, which could entail a waste of resources.

Similarly, high proportion of male students[8] and of those "affected" by diglossia do not contribute to improve the efficiency level of teachers. We observe that when a teacher shows imbalanced performances regarding the variables considered, the strong index attains low values and, as consequence, the mixed index decreases too. This implies that a teacher may have a weak index greater than another one, but the former may reach a mixed index value lower than the latter.

In this sense, particularly worrisome is the negative impact on the teaching-learning process of the index measuring the home resources for learning. This variable achieves very low –normalized– negative values for the least efficient teachers; therefore, we can conclude that having a low level of learning resources at the students' home has a negative effect for a teacher to be efficient, and this seems to be difficult to be compensated with good performances in other dimensions (at least, regarding the ones considered in our study). This effect is caused by the prevalence of high socio-economic

[7] Nevertheless, the data available do not allow measuring the quality of the weekly instruction time received by children in a precise manner for our purposes, so our results should be taken with caution in this area.

[8][39] studied risk determinants of school failure for Spain –using PISA 2006 data– and found that peer group characteristics like high proportion of girls at school decreased the likelihood of school failure.

inequality and the concentration of educational resources, and this lack of equal educational opportunities supposes a relevant loss of welfare to societies.

Additionally, a relevant conclusion is that "league tables" should be considered with caution, although they have been commonly used when international results are released [40]. Our results suggest that the development of such a ranking should not be only based on raw scores.

Acknowledgements

This work was supported by the Ministry of Economy, Industry and Competitiveness of Spain (ECO2017-88883-R and ECO2017-90573-REDT) and by the Andalusian Regional Ministry of Innovation, Science and Enterprises (PAI group SEJ-532). Ana B. Ruiz is recipient of a Post-Doctoral fellowship of "*Captación de Talento para la Investigación*" at the University of Malaga.

References

1. Hanushek, E.A., and Woessmann, L. (2007). *Education Quality and Economic Growth*. Washington, DC: World Bank.
2. Lockheed, M., and Hanushek, E. (1994). Concepts of educational efficiency and effectiveness. In: T. Husen and T.N. Postlethwaite, editors, *International encyclopedia of education,* Oxford, UK: Pergamon, p. 1779–1784.
3. Thompson-Schill, S.L., Ramscar, M., and Chrysikou, E.G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, 18: 259–263.
4. Wierzbicki, A.P., Makowski, M., and Wessels, J. (2000). *Model-Based Decision Support Methodology with Environmental Applications*. Dordrecht: Kluwer Academic Publishers.
5. Jürges, H., and Schneider, K. (2007). Fair ranking of teachers. *Empirical Economics*, 32: 411–431.
6. De Witte, K., Groot, W., and Maassen van den Brink, H. (2013). The efficiency of education in generating literacy: a stochastic frontier approach. *Review of Economics & Finance*, 2(2), 25–37.
7. Scippacercola, S., and D'Ambra, L. (2014). Estimating the relative efficiency of secondary schools by stochastic frontier analysis. *Procedia Economics and Finance*, 17: 79–88.
8. Lopez-Agudo, L.A., and Marcenaro-Gutierrez, O.D. (2017). Engaging children in lessons: the role of efficient and effective teachers. *School Effectiveness and School Improvement*, 28(4): 650–669.

9. Essid, H., Ouellette, P., and Vigeant, S. (2010). Measuring efficiency of Tunisian schools in the presence of quasi-fixed inputs: A bootstrap data envelopment analysis approach. *Economics of Education Review*, 29(4): 589–596.
10. Haelermans, C., and Ruggiero, J. (2013). Estimating technical and allocative efficiency in the public sector: A nonparametric analysis of Dutch schools. *European Journal of Operational Research*, 227(1): 174–181.
11. Blackburn, V., Brennan, S., and Ruggiero, J. (2014). Measuring efficiency in Australian schools: A preliminary analysis. *Socio-Economic Planning Sciences*, 48(1): 4–9.
12. Wanke, P., Blackburn, V., and Pestana-Barros, C. (2016). Cost and learning efficiency drivers in Australian schools: a two-stage network DEA approach. *Applied Economics*, 48(38): 3577–3604.
13. Charnes, A., Cooper, W.W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2: 429–444.
14. Liu, J.S., Lu, L.Y.Y., Lu, W., and Lin, B.J.Y. (2013). A survey of DEA applications. *Omega*, 41(5): 893–902.
15. De Witte, K., and Lopez-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, 68(4): 339–363.
16. Crespo-Cebada, E., Pedraja-Chaparro, F., and Santin, D. (2014). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *Journal of Productivity Analysis*, 41(1): 153–172.
17. Johnson, A.L., and Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research*, 221(1): 197–210.
18. Podinovski, V.V., Ismail, I., Bouzdine-Chameeva, T., and Zhang, W. (2014). Combining the assumptions of variable and constant returns to scale in the efficiency evaluation of secondary schools. *European Journal of Operational Research*, 239(2): 504–513.
19. Agasisti, T., and Johnes, G. (2015). Efficiency, costs, rankings and heterogeneity: The case of US higher education. *Studies in Higher Education*, 40(1): 60–82.
20. Reynolds, D. (2010). *Failure Free Education? The Past, Present and Future of School Effectiveness and School Improvement*. London: Routledge.
21. Braga, M., Paccagnella, M., and Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41: 71–88.
22. Creemers, B.P.M., and Reezigt, G.J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*, 7: 197–228.
23. 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*, 19: 429–446.
24. Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B., Timperley, H., and Earl, L. (2014). State of the art – Teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25: 231–256.

25. Ruiz, F., Cabello, J.M., and Luque, M. (2011). An application of reference point techniques to the calculation of synthetic sustainability indicators. *Journal of the Operational Research Society*, 62: 189–197.
26. Luque, M., Perez-Moreno, S., and Rodriguez, B. (2016). Measuring human development: A multi-criteria approach. *Social Indicators Research*, 125(3): 713–733.
27. Perez-Moreno, S. Luque, M., Rodriguez, B. (2016). Assessing global competitiveness under multi-criteria approach. *Economic Modelling*, 53(1): 398–408.
28. Luque, M., Perez-Moreno, S., Robles, J.A., Rodriguez, B. (2017). Measuring child and maternal health in developing countries: A proposal of new hybrid MDG composite indices. *Applied Research in Quality of Life*, 12(3): 737–758.
29. Ruiz, F., Cabello, J.M., Perez-Gladish, B. (2018). Building Ease-of-Doing-Business synthetic indicators using a double reference point approach. *Technological Forecasting and Social Change*, 131: 130–140
30. Ishizaka, A., Resce, G., and Mareschal, B. (2018). Visual management of performance with PROMETHEE productivity analysis. *Soft Computing*, 22(22): 7325–7338.
31. Martin, M.O., and Mullis, I.V.S. (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
32. OECD (2017). *PISA 2015 technical report*. Retrieved from http://www.oecd.org/pisa/data/2015-technical-report/.
33. Mullis, I.V.S., Martin, M.O., Foy, P., and Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
34. Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic Publishers.
35. Luque, M., Miettinen, K., Eskelinen, P., and Ruiz, F. (2009). Incorporating Preference Information in Interactive Reference Point Methods for Multiobjective Optimization. *Omega*, 37(2): 450–462.
36. Wierzbicki, A.P. (1980). The use of reference objectives in multiobjective optimization. In G. Fandel, and T. Gal, editors, *Multiple Criteria Decision Making, Theory and Applications*, Berlin: Springer-Verlag, p. 468–486.
37. Lopez-Agudo, L., and Marcenaro-Gutierrez, O.D. (2018). Are Spanish children taking advantage of their weekly classroom time? *Child Indicators Research*, in press, D.O.I. https://doi.org/10.1007/s12187-018-9537-4.
38. Gromada, A., and Shewbridge, C. (2016). Student learning time: A literature review. *OECD Education Working Papers,* No. 127. Paris: OECD Publishing.
39. Calero, J., Choi, A., and Waisgrais, S. (2010). Determinants of the school failure risk in Spain: a multilevel logistic model approach to PISA–2006. *Revista de Educación*, número extraordinario 2010: 225–256.

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