



Comparing the evolution of productivity and performance gaps in education systems through DEA: an application to Latin American countries

Juan Aparicio¹ · Sergio Perelman² · Daniel Santín³

Received: 5 June 2019 / Revised: 27 March 2020 / Accepted: 9 June 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The main objective of this paper is to propose a tool for measuring the productivity and performance gaps across a set of Decision Making Units for monitoring their evolution and analyzing their components over time. To do this, we use the approach proposed by Aparicio and Santín (Eur J Oper Res 267(1):227–235, 2018) which is grounded on a base-group base-period productivity index and Data Envelopment Analysis. Additionally, we propose a new index for measuring the performance gap between two or more groups of production units and its decomposition in effectiveness gap and outcome possibility set gap. As an empirical illustration of the approach, we focus our attention on the educational sector. In particular, we analyze six Latin American countries over time. For this purpose, we rely on OECD–PISA data aggregated at school level. Over the period 2006–2018, performance and productivity followed very different paths in each country showing that the correlation between school performance and productivity is very low. Therefore, we suggest that the simultaneous analysis of performance and productivity gaps together with their evolution over time is a must in order to benchmark countries and monitor improvements and weaknesses in education systems.

Keywords Data envelopment analysis · Performance · Productivity · Education

✉ Juan Aparicio
j.aparicio@umh.es

Sergio Perelman
sergio.perelman@ulg.ac.be

Daniel Santín
dsantin@ccee.ucm.es

¹ Center of Operations Research (CIO), Miguel Hernandez University of Elche, Av. de la Universidad s/n, 03202 Elche, Spain

² HEC-Management School, University of Liège, Bd Rectorat (B31), 4000 Liège, Belgium

³ Department of Applied Economics, Public Economics and Political Economy and Complutense Institute of Economic Analysis, Complutense University of Madrid, Campus de Somosaguas, 28223 Pozuelo de Alarcón, Spain

1 Introduction

The academic results of students measured using standardized tests scores at different educational levels are commonly used to measure and compare the quality of education systems and the future potential of human capital with respect to economic growth across countries. The main reason is that the number of quality-adjusted years of schooling calculated for a country has been shown to be strongly and positively correlated with greater economic growth rates (Hanushek and Kimko 2000; Hanushek and Woessmann 2008, 2011a, b). This link between educational achievement and growth is especially important across Latin American countries whose low-test scores have led to the lower economic growth of this area than in other world regions (Hanushek and Woessmann 2012). For example, as Kaarsen (2014) shows using TIMSS (Trends in Math and Science Study) data, one year of schooling in the United States is equivalent to more than one-and-a-half years of schooling in Chile or more than 2 years of schooling in Colombia.

In the light of this evidence, the evaluation of public educational policies and the monitoring of school performance are increasingly important for benchmarking the evolution of different education systems in one and the same regional context. For this purpose, this paper focuses on the measurement of two relevant economic dimensions—performance and productivity. In order to develop our approach, we will focus on the educational context, where both dimensions are central. With this objective, we define performance as a measure that captures the outcomes achieved in education (average student test scores in different subjects), regardless of the resources used to achieve these results. Additionally, two schools might be identical in terms of performance if they achieved exactly the same outcomes, whereas they may differ in terms of productivity if a different amount of resources were used to produce the same outcomes. Consequently, school productivity may correlate poorly or even negatively with school performance, as illustrated by Cherchye et al. (2019) for the case of secondary schools in the Netherlands.

Given that the educational production function includes many resources and outcomes, the measurement of performance and productivity is a problem of aggregating indicators into a single composite variable. Weighting can be done using different criteria. Principal component analysis or multi-criteria decision analysis, for example, are two approaches that could be used for this purpose (OECD 2005). In our case, we use production frontiers estimated by data envelopment analysis (DEA) to measure and compare performance and productivity gaps and their evolution over time. DEA can be used to build both dimensions calculating technical efficiency and effectiveness as distances to a production frontier. DEA has been widely applied in education (see De Witte and López-Torres 2017, for a review) to measure technical efficiency (Farrell 1957) because it does not impose a parametric functional form on the production function. Additionally, it is capable, with just a few assumptions, monotonicity and convexity, of handling multiple outputs and inputs. Therefore, each production unit has a free choice of weights to maximize its efficiency with respect to other units.

The application of DEA to define an effectiveness index through the aggregation of different outcomes was labeled as the benefit of the doubt (BoD) approach and was originally proposed by Melyn and Moesen (1991). The characterization provided by the BoD approach accounts for both the free choice of outcome weights and the fact that resources are not taken into account at all (in practice, a single constant equal to one resource is introduced in the analysis for all the observed units). The measurement of efficiency is directly related to the measurement of effectiveness. Keeping the same outcomes as for the measurement of effectiveness on the output side, efficiency also accounts for a set of resources on the input side using the DEA technique. In our case, the input selection reflects the educational production process, and the technical efficiency estimation generated by DEA reflects the maximum performance that can be achieved using the set of available resources, including school inputs and family background.¹

As mentioned above, education plays an important role in a global competitive economy, and it is necessary to evaluate the performance and productivity of a country and of a country compared with other countries, as well as monitoring gaps over time. Note, at this point, that results could be misleading if the focus is on just one dimension. For example, a country could be highly efficient compared with other countries. However, efficiency will be useless if the education system does not provide citizens with the minimum knowledge and skills needed to lead a decent life, that is, is not effective. In the same vein, a country could achieve an admissible performance level in terms of educational outcomes at the cost of higher spending levels than other countries with similar results. Therefore, educational performance could very well be driven not only by the inefficient use of allocated resources but also by resource constraints. In this case, there is room for an educational policy designed to achieve both objectives at the same time.

Performance and productivity gaps can be measured and compared across countries thanks to international databases like the OECD Program for International Student Assessment (PISA). Every 3 years, PISA provides a representative sample of 15-year-old students and their schools in the participating countries. PISA outcomes, measured through test scores in mathematics, reading and sciences are meaningful in the sense that their values can be interpreted as key skills levels, providing information about the percentage of students that acquire basic skills (OECD 2013). At the same time, PISA provides students and school background variables gathered through the same questionnaires across countries related to the resources available to produce education.

Consequently, PISA has made it possible to compare mainly technical efficiency through cross-country analysis using different approaches as reported by Afonso and St. Aubyn (2006), De Jorge and Santín (2010), Thieme et al. (2012), Aristovnik (2012), Deutsch et al. (2013), Aristovnik and Obadić (2014) or Cordero et al.

¹ Outcomes used to measure effectiveness are all average test scores. To measure efficiency, we take into account the same outcomes defined in the same way in order to ensure comparability, and resources are defined per capita (per student) or as a ratio. Throughout this paper efficiency and effectiveness measures are the bricks for calculating performance and productivity gaps.

(2018). However, there are very few empirical studies analyzing how performance and productivity gaps evolve over time for a set of repeated countries. There are two explanations for this point. Firstly, national education authorities fail to provide administrative panel databases at student and/or school level. Where national education databases are available, they are not comparable with other countries. Secondly, participant schools, and obviously students, differ from one PISA wave to another as a result of the random selection of participants in international studies. This limits the implementation of the standard Malmquist index, which is based on panel data.

In an effort to solve this problem, Aparicio et al. (2017) and Aparicio and Santín (2018) recently adapted the well-known Camanho and Dyson (2006) group performance index to analyze and explain productivity gaps through their decomposition into technology and efficiency gaps across different groups of production units over time. Their approach is grounded on a base-group base-period index that overcomes circularity problems.

The aim of this paper is to propose a new index for measuring performance gaps between two or more sets of production units over time together with its decomposition in effectiveness gap and outcome possibility set gap. As an empirical illustration of the approach, we demonstrate that it is worthwhile to analyze both dimensions, performance and productivity gaps, in the educational context as an early warning for potential gaps in future economic growth and as a way of benchmarking educational policies that work in a similar region. Moreover, we address the relationship between performance and productivity gap changes to understand how both dimensions evolve over time. To illustrate this approach, we use the PISA 2006, 2012 and 2018 waves for six Latin American countries—Argentina, Brazil, Chile, Colombia, Mexico and Uruguay—, using data from Portugal in 2006 as a reachable reference technology.

Our main results show that for the 2006–2018 period, Chile was the Latin American benchmark country in terms of performance change, whereas Colombia and Argentina (Uruguay) were the two countries that gained (lost) in both dimensions, performance and productivity. Brazil (Mexico) had a positive performance (productivity) change at the cost of productivity (performance).

The paper is organized as follows. Section 2 introduces the concept of distance function and its measurement using DEA. It also introduces the methodology for measuring group productivity gaps over time and the new results for measuring the performance gap between two or more groups of production units together with its decomposition. Section 3 briefly describes PISA data, especially the outputs and inputs included in the analysis. Section 4 reports the estimation and discussion of results. Finally, Sect. 5 provides the conclusions of this research, and some implications for educational policy makers and future research lines.

2 Methodology

In this section, we briefly introduce the notation and methods that we apply later in order to get the results associated with the comparison of the six Latin American countries regarding their PISA results.

First of all, let us introduce some notation to be used in the explanations linked to the methodology. Let us assume that we have observed n^{A^s} decision-making units (DMUs) in group A in period s , $s = t, t + 1$, which produce output $y^{A^s} \in R_+^q$ from input $x^{A^s} \in R_+^k$ and that we have also observed n^{B^s} DMUs in group B in period s , $s = t, t + 1$, which produce output $y^{B^s} \in R_+^q$ from input $x^{B^s} \in R_+^k$.² The DMUs operating in group A in period s are represented by their input–output vector as $(x_j^{A^s}, y_j^{A^s})$, $j = 1, \dots, n^{A^s}$. In the same way, $(x_j^{B^s}, y_j^{B^s})$ denotes the input–output vector of DMU j , $j = 1, \dots, n^{B^s}$, belonging to group B in period s . $D^{A^s}(x_j^{B^h}, y_j^{B^h}) = \inf\{\theta : (x_j^{B^h}, y_j^{B^h}/\theta) \in T^{A^s}\}$ represents the Shephard output distance function from observation $(x_j^{B^h}, y_j^{B^h})$ in group B in period h , $h = t, t + 1$, to the frontier of the technology of group A in period s , $s = t, t + 1$, T^{A^s} . A similar notation is used for the distance of a DMU in A with respect to the technology of group B and for the distance for a DMU that belongs to the same group as the reference technology.

For contemporaneous periods of time, i.e. when $h = s$, the reciprocal of the Shephard output distance function, $D^{A^s}(x_j^{B^s}, y_j^{B^s})^{-1}$, may be interpreted as technical efficiency in an output-oriented framework. Specifically, this value indicates how much $y_j^{B^s}$ must be increased, keeping constant $x_j^{B^s}$, in order to project point $(x_j^{B^s}, y_j^{B^s})$ onto the frontier of the technology, T^{A^s} , in this particular case. Generally, even for mixed periods, the Shephard output distance function measures the distance from the assessed point to the frontier of the reference technology.

There are different methods for estimating T^{A^s} and T^{B^s} , $s = t, t + 1$ in the literature. Some are parametric and require the production frontier to be characterized through a mathematical expression and a set of parameters that must be econometrically estimated from the data. Nevertheless, there are alternative data-driven approaches that do not require this specification and, accordingly, are considered non-parametric. We refer, in particular, to data envelopment analysis (DEA), originally proposed by Charnes et al. (1978). DEA consists of determining the smallest polyhedral set that envelops the observations and satisfies some properties, like, for example, free disposability. We use this technique in order to estimate the educational production function in our context since DEA is one of the most used methodologies for measuring efficiency in education (De Witte and López-Torres 2017).

In DEA, the Shephard output distance function $D^{A^s}(x_0^{B^h}, y_0^{B^h})$, $s, h = t, t + 1$, is estimated, assuming constant returns to scale (CRS), as the reciprocal of the optimal value for ϕ of the following linear programming model:

² Note that here we use the terms “outputs” and “inputs” to refer to “outcomes” and “resources”, respectively. The idea is to respect the traditional presentation in the efficiency literature and the generality of the methodology. As already mentioned in the empirical analysis, outcomes and resources are measured as per capita values and ratios.

$$\begin{aligned}
& \max_{\lambda, \phi} \quad \phi \\
& \text{s.t.} \quad \sum_{j=1}^{n^{As}} \lambda_j x_{ij}^{As} \leq x_{i0}^{Bh}, \quad i = 1, \dots, k \\
& \quad \quad \sum_{j=1}^{n^{As}} \lambda_j y_{rj}^{As} \geq \phi y_{r0}^{Bh}, \quad r = 1, \dots, q \\
& \quad \quad \lambda_j \geq 0, \quad j = 1, \dots, n^{As}
\end{aligned} \tag{1}$$

A value for $D^{As}(x_0^{Bh}, y_0^{Bh})$ that is equal to one indicates that point (x_0^{Bh}, y_0^{Bh}) is located on the frontier of the technology T^{As} . A value of less than one denotes that (x_0^{Bh}, y_0^{Bh}) is enveloped by the frontier, and, finally, a value that is greater than one specifies that the assessed point is outside the reference technology. Moreover, as already mentioned, the Shephard output distance function and its reciprocal, the optimal value of model (1), is directly related to the notion of technical efficiency when the evaluated observation and the reference technology belongs to the same period of time.³

This paper is concerned with the notion of efficiency but also with the concept of effectiveness when DMUs (schools in our empirical application) provide the population with services, namely education. The distinction between efficiency and effectiveness may be clarified as follows. Effectiveness can be defined as the ability to achieve goals, regardless of the amount of resources that might be needed for this purpose (Cherchye et al. 2019), whereas efficiency relates realized outputs to utilized inputs (see Prieto and Zoffio 2001).⁴

Very different terminologies, like, for example, the benefit of the doubt (BoD) approach (see Melyn and Moesen 1991, and, more recently, the method proposed by Cherchye et al. 2004), have been used to implement the notion of effectiveness in DEA. Generally speaking, the idea behind the measurement of effectiveness using DEA techniques consists of applying a model like (1), albeit omitting the observed inputs and using instead a single input equal to one for all the units (see also Lovell and Pastor 1999, who define this idea as a ‘pure output model’).

Under these premises, it is possible to define the Shephard output distance function to measure effectiveness in DEA through the reciprocal of the optimal value of the linear program (2).

³ Test scores in Latin America are below the average of OECD countries. For this reason, we consider that an output orientation is more appealing for policy makers interested in maximizing students’ achievements subject to their budget constraints.

⁴ For Mbuvi et al. (2012), effectiveness versus efficiency can be also loosely stated as “doing the right things” versus “doing things right”.

$$\begin{aligned}
 & \max_{\gamma, \phi_e} \quad \phi_e \\
 & \text{s.t.} \quad \sum_{j=1}^{n^{A^s}} \gamma_j \leq 1, \\
 & \quad \quad \sum_{j=1}^{n^{A^s}} \gamma_j y_{rj}^{A^s} \geq \phi_e y_{r0}^{B^h}, \quad r = 1, \dots, q \\
 & \quad \quad \gamma_j \geq 0, \quad j = 1, \dots, n^{A^s}
 \end{aligned} \tag{2}$$

The Shephard output distance function $D^{A^s}(1, y_0^{B^h})$ and its reciprocal, in this case the optimal value of model (2) ϕ_e , is the measure of effectiveness.

Additionally, it is of interest in some situations, such as the one described in our empirical study, to extend the analysis in order to compare the productivity and the performance of several natural groups of DMUs.⁵ To make these comparisons, Camanho and Dyson (2006) introduced a Malmquist-type index (hereinafter referred to as CDMI) to provide an average indicator of the relative productivity gap of two groups of DMUs within a period of time.

2.1 Camanho–Dyson Malmquist index approach

The CDMI by Camanho and Dyson (2006) is an adaptation of the Malmquist productivity index to provide a cross-sectional comparison of the productivity of two groups of DMUs operating under different conditions and that, therefore, can correspond to different technologies. Given two groups of units for comparison, the adaptation of the Malmquist productivity index basically consists of using the units of one group as observations of period t and treating the DMUs of the other group as observations corresponding to period $t + 1$. The CDMI for comparing the productivity of two groups of DMUs, A and B, associated with different programs, ownership types or education systems in one time period s is defined as follows:

$$\text{CDMI}_s^{AB} = \left[\frac{\left(\prod_{j=1}^{n^{A^s}} D^{A^s}(x_j^{A^s}, y_j^{A^s}) \right)^{1/n^{A^s}}}{\left(\prod_{i=1}^{n^{B^s}} D^{A^s}(x_i^{B^s}, y_i^{B^s}) \right)^{1/n^{B^s}}} \cdot \frac{\left(\prod_{j=1}^{n^{A^s}} D^{B^s}(x_j^{A^s}, y_j^{A^s}) \right)^{1/n^{A^s}}}{\left(\prod_{i=1}^{n^{B^s}} D^{B^s}(x_i^{B^s}, y_i^{B^s}) \right)^{1/n^{B^s}}} \right]^{1/2} \tag{3}$$

where $(\prod_{i=1}^{n^{B^s}} D^{A^s}(x_i^{B^s}, y_i^{B^s}))^{1/n^{B^s}}$, is the average distance of DMUs in group B relative to group A technology. The CDMI is straightforward to interpret, where a value higher (lower) than one indicates better (worse) average productivity in group A

⁵ In education, it is meaningful to compare the productivity and the performance of groups of schools in different environments, as well as different regions inside the same country, different countries with similar economic features (as in this study) or with different regulatory conditions, according to school ownership (public vs. private), religion (catholic vs. non-catholic) or location (urban vs. rural), for example. Recently, Amado et al. (2018) compare the gender pay gap using a similar approach.

than in group B. The relative productivity gap measured in (3) can be decomposed into the following terms:

$$CDMI_s^{AB} = \underbrace{\frac{\left(\prod_{j=1}^{n^A} D^A(x_j^A, y_j^A)\right)^{1/n^A}}{\left(\prod_{i=1}^{n^B} D^B(x_i^B, y_i^B)\right)^{1/n^B}}}_{EG_s^{AB}} \cdot \underbrace{\left[\frac{\left(\prod_{i=1}^{n^A} D^B(x_i^A, y_i^A)\right)^{1/n^A}}{\left(\prod_{j=1}^{n^B} D^A(x_j^B, y_j^B)\right)^{1/n^B}} \cdot \frac{\left(\prod_{i=1}^{n^B} D^B(x_i^B, y_i^B)\right)^{1/n^B}}{\left(\prod_{i=1}^{n^A} D^A(x_i^A, y_i^A)\right)^{1/n^A}} \right]^{1/2}}_{TG_s^{AB}} \quad (4)$$

where EG_s^{AB} compares within-group efficiency spreads, measuring the technical efficiency gap between both groups, while the ratio TG_s^{AB} evaluates the technology gap between the frontiers of the two analyzed groups, A and B. EG_s^{AB} and TG_s^{AB} are interpreted like the CDMI. When $EG_s^{AB} > 1$, average technical efficiency in group A is higher than in group B (and vice versa when $EG_s^{AB} < 1$), which means that the DMUs in group A are on average closer to their production frontier than their counterparts in group B. When $TG_s^{AB} > 1$, the production technology in group A dominates the frontier in group B, and more outputs can be produced with the same quantities of inputs (and vice versa when $TG_s^{AB} < 1$).

As Camanho and Dyson (2006) claimed, the main advantage of this approach is probably that it does not assume convex combinations of units belonging to different groups under different circumstances to be feasible. This means that there is no mixing of technologies, as opposed to other alternatives that are based upon pooling all DMUs together to form a common merged synthetic frontier or meta-frontier. Thanks to this interesting feature, it has been possible to successfully use the CDMI in empirical applications to provide comparisons of groups of DMUs in cross-sectional studies (see, for example, Vaz and Camanho 2012, Ferreira and Marques 2015, Thanassoulis et al. 2015 and Amado et al. 2019).

The CDMI was recently extended by Aparicio et al. (2017) and Aparicio and Santín (2018) for use in the context where panel or pseudo-panel data are available, and the target is to determine how the measured productivity gap initially measured by the CDMI changes over time. A panel implies that the same group of DMUs is observed across all analyzed periods. However, PISA is based on the selection of random waves of representative samples of students and schools at regional or country level in order to analyze their test scores over time. The schools contained in each wave vary from one period to another and are usually anonymous for researchers. Aparicio et al. (2017) refer to this information structure as a pseudo-panel database. In this case, the approach introduced in Aparicio et al. (2017) and Aparicio and Santín (2018) can be used to compare the change in the productivity gap of these representative groups of DMUs over time using a pseudo-panel Malmquist-type index (PPMI).

In this paper, we use the approach proposed by Aparicio and Santín (2018) rather than by Aparicio et al. (2017) to compare countries over time because it solves two open questions that arose in Aparicio et al. (2017). The first one is related to the circularity property, a problem also faced by Camanho and Dyson (2006), when comparing more than two groups of units, while the second one is the possibility of

interpreting the PPMI in terms of the ratio of aggregated productivity changes in the two evaluated groups between periods t and $t + 1$.

2.2 Aparicio–Santín Malmquist index approach

Let us now introduce the groundwork of the approach proposed by Aparicio and Santín (2018), which is based on the standard Shephard output distance function to deal with the comparison of two or more groups of DMUs under the notion of efficiency.⁶

In an attempt to improve the properties of the original CDMI, Aparicio and Santín (2018) introduced a new base-group CDMI. Let us assume that, apart from the relevant groups A and B that we want to compare, we have observed a set of DMUs of a so-called reference group R. Then, the base-group CDMI for comparing the productivity of groups A and B in a single time period s with respect to R in period h , $CDMI_s^{AB}(R^h)$, is defined as follows:

$$CDMI_s^{AB}(R^h) = \frac{\left(\prod_{j=1}^{n^{As}} D^{R^h}(x_j^{As}, y_j^{As})\right)^{1/n^{As}}}{\left(\prod_{i=1}^{n^{Bs}} D^{R^h}(x_i^{Bs}, y_i^{Bs})\right)^{1/n^{Bs}}}. \tag{5}$$

Note that all Shephard output distance functions in (5) are determined taking T^{R^s} as the reference technology. Additionally, it is straightforward to prove that, assuming the simplest case, i.e. only one input and only one output, $CDMI_s^{AB}(R^h) = \left(\prod_{j=1}^{n^{As}} \frac{y_j^{As}}{x_j^{As}}\right)^{1/n^{As}} / \left(\prod_{i=1}^{n^{Bs}} \frac{y_i^{Bs}}{x_i^{Bs}}\right)^{1/n^{Bs}}$. In other words, the base-group CDMI compares the average productivities in group A with the average productivities in group B. This is a consequence of assuming CRS in the computation of each Shephard output distance function. This is a must if, as in our case, the Malmquist productivity index is used.⁷ This simple result justifies our claim that the base-group CDMI really measures the productivity gap between groups A and B. The interpretation in numerical values is exactly the same as the standard CDMI. When $CDMI_s^{AB}(R^s) > 1$, the average productivity of production units in group A is higher than in group B. Also, as in (3), (5) can be decomposed into two subcomponents:

$$CDMI_s^{AB}(R^h) = \underbrace{\frac{\left(\prod_{j=1}^{n^{As}} D^{As}(x_j^{As}, y_j^{As})\right)^{1/n^{As}}}{\left(\prod_{i=1}^{n^{Bs}} D^{Bs}(x_i^{Bs}, y_i^{Bs})\right)^{1/n^{Bs}}}}_{EG_s^{AB}} \cdot \underbrace{\left[\frac{\left(\prod_{j=1}^{n^{As}} D^{R^h}(x_j^{As}, y_j^{As})\right)^{1/n^{As}}}{\left(\prod_{j=1}^{n^{As}} D^{As}(x_j^{As}, y_j^{As})\right)^{1/n^{As}}} \cdot \frac{\left(\prod_{i=1}^{n^{Bs}} D^{Bs}(x_i^{Bs}, y_i^{Bs})\right)^{1/n^{Bs}}}{\left(\prod_{i=1}^{n^{Bs}} D^{R^h}(x_i^{Bs}, y_i^{Bs})\right)^{1/n^{Bs}}} \right]}_{TG_s^{AB}(R^h)} \tag{6}$$

⁶ Note at this point that all the indexes introduced in this section calculate productivity gaps and their evolution with respect to the efficiency dimension, i.e. including inputs and outputs for measuring distances. The same indexes for the effectiveness dimension are defined and explained in Sect. 2.3.

⁷ In this paper we assume a CRS technology for avoiding the infeasibility issue (Lovell, 2003). Moreover, in our empirical application scale is not a major issue because education is produced at classroom level and outputs and inputs variables are usually averaged or normalized for avoiding size issues.

The first components in (4) and (6) are identical, whereas the second terms, associated with the comparison of the frontiers of each group, are different, although their numerical interpretation in terms of efficiency gap and productivity gap are the same as in (4). The technology gap $TG_s^{AB}(R^h)$ in (6) satisfies the circularity property, which is an additional advantage of assuming a reference group to define the base-group CDMI. This property indicates that, if more than two groups are taken into account, for example A, B and C, the index for comparing A to C is equal to the product of the index comparing A to B and the index comparing B to C. Like the base-period Malmquist index, which passes the circularity test (Berg et al. 1992), the base-group CDMI and its components are also circular.

In order to deal with panels and pseudo-panels, which is actually the context with respect to PISA data, Aparicio and Santín (2018) suggest using a base-group base-period PPMI to measure the relative productivity gap change between A and B within t and $t + 1$:

$$PPMI_{t,t+1}^{AB}(R^h) = \frac{CDMI_{t+1}^{AB}(R^h)}{CDMI_t^{AB}(R^h)}, \tag{7}$$

where (7) is the ratio of two base-group CDMIs. To calculate $PPMI_{t,t+1}^{AB}(R^h)$, the reference group and the period of time for evaluating the base-group CDMI at t and $t + 1$ are both fixed.

One of the advantages of the PPMI proposed by Aparicio and Santín is that the index can be interpreted as the ratio between an aggregated measure of productivity changes, from t to $t + 1$, of the units corresponding to group A and an aggregated measure of productivity changes, from t to $t + 1$, of the units belonging to B. To illustrate this point, let us suppose that we have real panel data, then $n^{A'} = n^{A^{t+1}} = n^A$ and $n^{B'} = n^{B^{t+1}} = n^B$. In this case, the $PPMI_{t,t+1}^{AB}(R^h)$ can be rewritten as:

$$PPMI_{t,t+1}^{AB}(R^h) = \frac{\left(\prod_{j=1}^{n^A} M^{R^h}(x_j^{A^{t+1}}, y_j^{A^{t+1}}, x_j^{A^t}, y_j^{A^t})\right)^{1/n^A}}{\left(\prod_{i=1}^{n^B} M^{R^h}(x_i^{B^{t+1}}, y_i^{B^{t+1}}, x_i^{B^t}, y_i^{B^t})\right)^{1/n^B}}, \tag{8}$$

where $M^{R^h}(x_j^{A^{t+1}}, y_j^{A^{t+1}}, x_j^{A^t}, y_j^{A^t})$ and $M^{R^h}(x_i^{B^{t+1}}, y_i^{B^{t+1}}, x_i^{B^t}, y_i^{B^t})$ denote the traditional Malmquist index for units of group A and units of group B, respectively, when the reference technology of in period h used is R .

The components of the PPMI can be interpreted similarly if it is decomposed in the usual terms. In this way, the efficiency gap change (EGC) can be interpreted as the ratio between an aggregated measure of efficiency changes, from t to $t + 1$, of the units in A and an aggregated measure of efficiency changes, from t to $t + 1$, of the units belonging to B, whereas the technology gap change (TGC) can be regarded as the ratio of the aggregated technical change in group A to the aggregated technical change in group B. See the specific definitions of each term in (9).

$$PPMI_{t,t+1}^{AB}(R^h) = \frac{CDMI_{t+1}^{AB}(R^h)}{CDMI_t^{AB}(R^h)} = \frac{EG_{t+1}^{AB}}{EG_t^{AB}} \cdot \frac{TG_{t+1}^{AB}(R^h)}{TG_t^{AB}(R^h)} = EGC_{t,t+1}^{AB} \cdot TGC_{t,t+1}^{AB}(R^h). \tag{9}$$

As for the base-group CDMI, the base-group base-period PPMI, and its components in (9) also satisfy the circular relation.

Finally, we can slightly modify (5) using this framework to measure the productivity changes of the same group of production units $PPMI_{t,t+1}^{AA}(R^h)$ over time as follows:

$$PPMI_{t,t+1}^{AA}(R^h) = \frac{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{R^h}(x_j^{A^{t+1}}, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}}{\left(\prod_{i=1}^{n^{A^t}} D^{R^h}(x_i^{A^t}, y_i^{A^t})\right)^{1/n^{A^t}}}. \tag{10}$$

In the same vein, (10) can be decomposed into two subcomponents:

$$PPMI_{t,t+1}^{AA}(R^h) = \underbrace{\frac{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{A^{t+1}}(x_j^{A^{t+1}}, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}}{\left(\prod_{i=1}^{n^{A^t}} D^{A^t}(x_i^{A^t}, y_i^{A^t})\right)^{1/n^{A^t}}}}_{EGC_{t,t+1}^{AA}} \cdot \underbrace{\frac{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{R^h}(x_j^{A^{t+1}}, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}}{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{A^{t+1}}(x_j^{A^{t+1}}, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}} \cdot \frac{\left(\prod_{i=1}^{n^{A^t}} D^{A^t}(x_i^{A^t}, y_i^{A^t})\right)^{1/n^{A^t}}}{\left(\prod_{i=1}^{n^{A^t}} D^{R^h}(x_i^{A^t}, y_i^{A^t})\right)^{1/n^{A^t}}}}_{TGC_{t,t+1}^{AA}(R^h)} \tag{11}$$

where $PPMI_{t,t+1}^{AA}(R^h)$ and its components $EGC_{t,t+1}^{AA}$ and $TGC_{t,t+1}^{AA}(R^h)$ also satisfy the circularity property.

2.3 New methodological results: The performance gap index and its decomposition

By analogy with the base-group CDMI and the PPMI defined by means of the standard Shephard output distance function, we extend these indexes in this paper to compare a number of groups of DMUs over time from the viewpoint of performance. To do this, we use just one input equal to one for all production units included in the analysis. In this way, the base-group performance gap index (PGI) can be defined from (5) as follows:

$$PGI_s^{AB}(R^h) = \frac{\left(\prod_{j=1}^{n^{A^s}} D^{R^h}(1, y_j^{A^s})\right)^{1/n^{A^s}}}{\left(\prod_{i=1}^{n^{B^s}} D^{R^h}(1, y_i^{B^s})\right)^{1/n^{B^s}}}. \tag{12}$$

If performance is defined as only one output, we can, accordingly, simplify (12) to $PGI_s^{AB}(R^h) = \left(\prod_{j=1}^{n^{A^s}} y_j^{A^s}\right)^{1/n^{A^s}} / \left(\prod_{i=1}^{n^{B^s}} y_i^{B^s}\right)^{1/n^{B^s}}$. Therefore, PGI no longer compares productivities of group A units with respect to productivities of group B units. Instead, it compares the aggregation of the outcomes of both groups regardless of the resources utilized to achieve the outcomes. For PGI, therefore, we speak in terms of an outcome gap rather than a productivity gap between groups as we do for the CDMI. Obviously, it is straightforward to demonstrate that the PGI can again be decomposed, like the base-group CDMI in (6), into two terms, which, for the sake of

clarity, we will refer to as effectiveness gap and outcome possibility set gap, as follows:

$$PGI_s^{AB}(R^h) = \underbrace{\frac{\left(\prod_{j=1}^{n^{As}} D^{As}(1, y_j^{As})\right)^{1/n^{As}}}{\left(\prod_{i=1}^{n^{Bs}} D^{Bs}(1, y_i^{Bs})\right)^{1/n^{Bs}}}}_{Efg_s^{AB}} \cdot \underbrace{\left[\frac{\left(\prod_{i=j}^{n^{As}} D^{R^h}(1, y_j^{As})\right)^{1/n^{As}}}{\left(\prod_{j=1}^{n^{As}} D^{As}(1, y_j^{As})\right)^{1/n^{As}}} \cdot \frac{\left(\prod_{i=1}^{n^{Bs}} D^{Bs}(1, y_i^{Bs})\right)^{1/n^{Bs}}}{\left(\prod_{i=1}^{n^{Bs}} D^{R^h}(1, y_i^{Bs})\right)^{1/n^{Bs}}} \right]}_{OPSG_s^{AB}(R^h)}, \tag{13}$$

where Efg_s^{AB} is the effectiveness gap that represents the average distance of production units in group A to their outcome possibility set with respect to the average distance of production units in group B to their own outcome possibility set, and $OPSG_s^{AB}(R^h)$ represents the outcomes possibility sets gap between groups A and B, respectively. For both components, values greater (less) than one mean that group A (group B) performs better than group B (group A).

When dealing with panels and pseudo-panels, we define the base-group base-period performance gap change index to measure the performance gap index change PGIC between A and B within t and $t + 1$ as follows:

$$PGIC_{t,t+1}^{AB}(R^h) = \frac{PGI_{t+1}^{AB}(R^h)}{PGI_t^{AB}(R^h)} = \frac{Efg_{t+1}^{AB}}{Efg_t^{AB}} \cdot \frac{OPSG_{t+1}^{AB}(R^h)}{OPSG_t^{AB}(R^h)} = EfgC_{t,t+1}^{AB} \cdot OPSGC_{t,t+1}^{AB}(R^h). \tag{14}$$

Again the PGIC can, like the PPMI in (9), be straightforwardly decomposed into effectiveness gap change $EfgC_{t,t+1}^{AB}$ and outcome production set gap change $OPSGC_{t,t+1}^{AB}(R^h)$. The $PGIC_{t,t+1}^{AA}(R^h)$ can be easily calculated and decomposed into two time periods for the same group of production units following (10) as follows.

$$PGIC_{t,t+1}^{AA}(R^h) = \frac{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{R^h}(1, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}}{\left(\prod_{i=1}^{n^{A^t}} D^{R^h}(1, y_i^{A^t})\right)^{1/n^{A^t}}}. \tag{15}$$

Finally, (15) can be decomposed into two subcomponents following (11) as:

$$PGIC_{t,t+1}^{AA}(R^h) = \underbrace{\frac{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{A^{t+1}}(1, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}}{\left(\prod_{i=1}^{n^{A^t}} D^{A^t}(1, y_i^{A^t})\right)^{1/n^{A^t}}}}_{EjGC_{t,t+1}^{AA}} \cdot \underbrace{\left[\frac{\left(\prod_{i=j}^{n^{A^{t+1}}} D^{R^h}(1, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}}{\left(\prod_{j=1}^{n^{A^{t+1}}} D^{A^{t+1}}(1, y_j^{A^{t+1}})\right)^{1/n^{A^{t+1}}}} \cdot \frac{\left(\prod_{i=1}^{n^{A^t}} D^{A^t}(1, y_i^{A^t})\right)^{1/n^{A^t}}}{\left(\prod_{i=1}^{n^{A^t}} D^{R^h}(1, y_i^{A^t})\right)^{1/n^{A^t}}} \right]}_{OPSG_{t,t+1}^{AA}(R^h)} \tag{16}$$

3 Data

We selected PISA for this benchmark study because it contains the best suited information for measuring performance and productivity gaps for different groups of schools. On the output side, the average scores gathered in the standardized tests taken by 15-year-old students from the same school are a good measure of expected educational outcomes.⁸ On the input side, the information available on school inputs and family background are very close to the definition of educational resources to be accounted for by an ideal model.

In this paper, we compare the performance and productivity dimensions for the educational production of six Latin American countries participating in the PISA 2006, 2012 and 2018 waves: Argentina, Brazil, Chile, Colombia, Mexico, and Uruguay. For this purpose, we use student data from the three waves aggregated at school level.⁹

As mentioned above, a reference sample is necessary to compute the PPMI and the PGIC to assure circularity. There is no simple rule for selecting this reference technology. The technology of some DMU groups evaluated in a particular time period could be selected as the reference, although the choice between the evaluated groups is likely to be arbitrary, as there is no reason to choose one group over the others. To avoid this issue, we prefer to select a group of DMUs outside the evaluated groups with a technology that is reachable for all of them.¹⁰ For this purpose, we selected the technology defined by the PISA 2006 Portuguese schools sample as the baseline. As the descriptive statistics reported later in this section show, the characteristics of this Portuguese sample are in most respects comparable to the Latin American countries samples for 2006, 2012 and 2018, and it is a reachable benchmark for this set of countries.

From the original PISA sample, we picked schools with complete data for the selected variables. We removed schools considered as outliers whose values were three standard deviations away from the average values within each country. Table 1 reports the composition of the final sample by wave and country. We also included data from the 2012 and 2018 Portuguese sample for comparative purposes, as shown later. The inputs and outputs choice for our analysis relies on the well-known educational production function proposed by Levin (1974) where a vector of inputs can produce a vector of outputs assuming the existence of inefficient behaviors. Under this framework, schools are evaluated according to their ability of making the most with their inputs.

⁸ Another key dimension of countries' educational outcomes is children's enrollment rates at secondary schools. Unfortunately, this dimension cannot be considered in this study given that the information is not available at school level in PISA.

⁹ Two other Latin American countries, Costa Rica and Peru, participated in PISA 2012 and 2018, but not in PISA 2006. This is the reason why they were not included in this analysis. In 2021, four other Latin American countries will join the PISA project: El Salvador, Guatemala, Jamaica and Paraguay.

¹⁰ Ideally, the reference should be chosen by agreement by all evaluated groups. As a rule of thumb, the reference should be recognized as a medium-term target with comparable sample size and as similar as possible in culture and legal framework.

3.1 Outputs

We selected the average results at school level in reading, mathematics and science as outcomes. It is worth underlying that there are different cognitive and non-cognitive dimensions embedded in the educational outcomes received by an individual; what people usually calls a ‘good education’. Nevertheless, from the economics of education point of view, the focus is currently set in measuring students’ achievements through standardized tests. As it was remarked in the introduction, there is a strong evidence in the literature for considering that the quality of education, measured as the average results on test scores in mathematics, science and reading using international tests like PISA, has a highly statistically significant effect on the growth of real GDP per capita (see Hanushek and Woessmann 2020 for a discussion).

Note here that results in different PISA waves are only directly comparable for some years and subjects (OECD 2014a, p. 52–53). As OECD (2014b, p. 159) reports; “Again, for PISA 2012 the decision was made to report the reading, mathematics and science scores on these previously developed scales. That is the reading scales used for PISA 2000, PISA 2003, PISA 2006, PISA 2009 and PISA 2012 are directly comparable. PISA 2012 mathematics reporting scale is directly comparable to PISA 2003, PISA 2006 and PISA 2009 and the science reporting scale is directly comparable to PISA 2006 and PISA 2009 scale”. This means that PISA scores in the three subjects are only comparable from 2006 onwards. This led us to select PISA 2006, 2012 and 2018 and exclude earlier waves.

Table 1 also reports the average student scores and percentage variations between the 2006, 2012 and 2018 waves for Latin American countries and Portugal that constitute our three outputs. Compared with 15-year-old Portuguese students, Latin American students achieve, on average, lower scores for the three periods and for all three subjects, ranging from 20 to nearly 100 points. This is a large gap by PISA standards. Furthermore, scores for several countries and subjects were below 400 points in 2006 and still below 400 points in 2012 and 2018. However, we find that most countries improved average scores substantially from 2006 to 2018. This is the case of Chile, Colombia, Argentina and, Brazil. On the contrary, Mexican and particularly Uruguayan students’ scores declined. Mexican and Uruguayan students, which were among the best Latin American performers in 2006, were outperformed by Chilean students in 2018.

3.2 Inputs

On the input side, we separate school resources from family background for producing educational outcomes, as suggested by Hanushek et al. (2013). A drawback of PISA databases is that some resource scores provided by PISA, like economic, social and cultural status (*ESCS*) and school educational resources (*SCMATEDU* or *SCMATBUI*) are composite indexes built by categorical principal component analysis and centered at zero in each wave. This means that these numbers cannot be interpreted in absolute terms to measure productivity progress over time from one wave to another. Therefore, we have to be cautious about the selection of the inputs that can be used to interpret

Table 1 School sample and average PISA student scores in mathematics, reading and science

Country	Wave	School sample	Mathematics		Reading		Science	
			Score	$\Delta\%$	Score	$\Delta\%$	Score	$\Delta\%$
Argentina	2006	159	378		372		388	
	2012	212	390	3.3	396	6.3	405	4.3
	2018	381	384	-1.6	407	2.8	409	1.0
Brazil	2006	579	358		382		378	
	2012	812	369	3.0	385	0.8	382	1.1
	2018	504	377	2.3	406	5.4	397	3.9
Chile	2006	149	396		424		422	
	2012	204	434	9.6	449	6.1	455	7.7
	2018	198	425	-2.1	460	2.4	452	-0.6
Colombia	2006	142	369		384		388	
	2012	303	381	3.5	408	6.2	403	3.9
	2018	237	395	3.5	415	1.9	417	3.5
México	2006	1050	417		424		419	
	2012	1346	412	-1.2	421	-0.7	413	-1.4
	2018	277	406	-1.5	417	-0.9	416	0.7
Uruguay	2006	265	430		417		431	
	2012	173	404	-6.0	406	-2.5	411	-4.6
	2018	183	408	1.0	415	2.3	416	1.2
<i>Reference country</i>								
Portugal	2006	168	465		471		474	
	2012	165	483	4.0	484	2.7	486	2.5
	2018	235	484	0.0	481	-0.5	483	-0.5

results as absolute values over time, which is our purpose. Another drawback is that PISA does not report information on school accounts. Therefore, it is not possible to make comparisons in monetary terms, e.g. cost reduction. Nevertheless, the variables representing school resources are, to some extent, proxies of school costs.

In our model specification, school resources are represented by two inputs: the *teacher/student ratio* and a *school resources index*. Both variables are computed at school level and rely on information collected by PISA with a specific survey questionnaire answered by school principals. The first input is the *teacher/student ratio* built as the ratio of the total number of teachers weighted by their working hours (part-time teachers contribute 0.5 and full-time teachers 1) to the total number of pupils. It approximates the labor factor.

The second input is the school resource index that is built adding up the responses to several items related to school deficiencies. Following Santín and Sicilia (2015), we select matching items across waves 2006, 2012 and 2018 and give each school one point for every item for which the principal's response was 'not deficient'. Examples of items for PISA 2006 and 2012 included in question SC14 are: 'Qualified science teachers', 'Qualified mathematics teachers', 'Qualified reading teachers', 'Any other personal support', 'Science laboratory equipment', 'Educational material', 'Computers',

‘Software’, ‘Library resources’, ‘Audiovisual resources’. For comparability with the two previous waves, in PISA 2018 we weighted the responses included in question SC017 (teaching staff, assisting staff, educational material and the quality of educational material and question SC155 (digital devices and software). This variable was rescaled, and, therefore, the minimum value is one and the maximum eleven in order to avoid zero values in the empirical analysis. This variable approximates the capital factor.

As student resources, we use two inputs, which, averaged over students, represent students’ family background at the school level representing the raw material for producing education. Therefore, the third input is the *highest number of years of schooling of parental education (PARED)* that allows international comparability. According to the International Standard Classification of Education (ISCED; OECD 1999), the student’s mother and the student’s father educational levels are transformed into an estimated number of years of schooling. The higher value of either parent or the only available parent information is assigned to each student. Finally, the fourth input is the *index of the highest parental occupational status (HISEI)*, coded according to International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom et al. 1992). Using open-ended questions about occupational data of both parents the answers were mapped into the ISEI for the student’s mother and the student’s father. The highest occupational status of parents (HISEI) corresponds to the higher ISEI score of either parent or to the only available parent’s ISEI score (for more details about these two inputs see OECD 2014b). Note at this point that we have selected standard variables that are consistent with previous research (De Witte and López-Torres 2017).

Table 2 reports the average values of school and family background resources. Surprisingly, there are big differences in the teacher/student ratio across Latin American countries. In 2018, for instance, this ratio was three times higher in Argentina (13.52) than in Brazil (4.57) and equivalent to Portugal (11.51). As expected, differences across countries in the average school resources index are smaller. Both the teacher/student ratio and the resources index values increased from 2006 to 2018 across all countries, with the exception of Colombia, whose teacher/student ratio diminished from 5.25 to 4.82, while school resources remained stable showing just a slight decline in Argentina from 7.57 to 6.72.

Interestingly, student resources show some variation across countries and across waves. Chile is the country with the highest values in 2018 for both variables—higher parental education and occupation—, and, together with Colombia, the only country for which these two indicators increased, on average, from 2006 to 2018. There are negative variations in both indicators for Uruguay, and in parental socio-economic status for three others, Argentina, Brazil and Mexico. The general better parental education can be explained by the increase in secondary education enrolment rates among 15-year-old students observed in most of these countries, with the exception of Chile.¹¹

¹¹ World Development Indicators from the World Bank (2020) report the secondary education enrollment evolution for the period 2006 to 2018 in these countries, as follows: 79.1% to 89.5% (2016) in Argentina; 73.2% (2007) to 81.7% (2017) in Brazil; 89.9% to 86.8% (2017) in Chile; 68.9% to 77.4% in Colombia; 67.8% to 81.2% (2017) in Mexico; and 67.6% (2007) to 88.2% (2017) in Uruguay. In Portugal, the reference country, the evolution over the same period was from 82.6% to 94.7%.

4 Base-group CDMI and the PGI

In this section, we report the cross-sectional Camanho-Dyson Malmquist indexes of productivity, $CDMI_s^{AB}(R^h)$ and performance, $PGI_s^{AB}(R^h)$, for each years and for each pair AB of Latin-American countries applying, respectively, Eqs. (6) and (13). R^h indicates the reference technology, which in our application corresponds to Portugal in $h=2006$. Note that two types of average school distances, by country and year, enter in the computation of Eqs. (6) and (13). Therefore, we compute distances with respect to the technology defined within the country sample and distances with respect to the reference benchmark technology, which corresponds to the Portuguese 2006 schools sample.

Then, first of all, we report in Table 3 the averages of DEA effectiveness and efficiency scores, computed using models 1 and 2. The first two columns (country reference) report the scores computed separately by wave and by country, assuming in each case a country-specific benchmark. This benchmark is composed of the schools retained in the sample for the respective country in the respective wave. We learn from the above results that school effectiveness scores lie on average between 63.5% (Brazil, 2006) and 80.9% (Chile, 2018) in Latin American countries and that, in most cases, these scores are on average lower than efficiency scores. The only exceptions are Mexico in the three waves and Brazil in 2012 and Colombia in 2018.¹²

The results reported in the following two columns of Table 3 account for the reference technology, namely, the 2006 Portuguese schools sample. They give a rather different view than the findings assuming country—and wave-specific technology. Average effectiveness scores are lower than the Portuguese benchmark in all Latin American countries. They vary from 65.1% (Brazil, 2006) to 76.7% (Chile, 2018), and they increased over the period in all cases, except in Mexico and Uruguay. These results mean that the test scores for Latin American countries in 2018 need to improve from 23 to 32% to reach Portuguese technology effectiveness. On the contrary, efficiency scores among Latin American countries are mostly higher than the reference used as the benchmark. The only exceptions are Argentina, and, to a lesser extent, Uruguay, in 2012. Nevertheless, we observe a decline in efficiency over the three waves in Brazil, Chile and Uruguay.

The last column in Table 3 reports Spearman's rank correlation coefficients between school effectiveness and efficiency scores within each country and the PISA wave using Portugal 2006 as the reference technology. As expected, correlations are very low and, in some cases, even negative. The correlation between school effectiveness and efficiency was positive and in the three waves just in Argentina and Colombia together with Portugal.

These results show per se the interest of designing educational policies to measure and compare effectiveness and efficiency simultaneously in the same school samples when consistent and complete data on outcome and resources are available. Schools performing very well in terms of effectiveness but not so well in terms of efficiency

¹² Note that this is also the case for Portugal in 2012 and 2018.

Table 2 Average school and family inputs in PISA by country and year

Country	Year	Teacher student (100)		School resources		Parental education		Parental socio-econ. status	
		Ratio	$\Delta\%$	Index	$\Delta\%$	Ratio	$\Delta\%$	Index	$\Delta\%$
Argentina	2006	9.79		7.57		11.90		44.9	
	2012	12.93	32.1	7.75	2.3	12.56	5.6	43.9	-2.1
	2018	13.52	4.5	6.72	-13.2	13.02	3.7	44.1	0.3
Brazil	2006	3.57		5.95		10.36		41.6	
	2012	4.06	13.6	6.62	11.3	9.68	-6.6	39.7	-4.5
	2018	4.57	12.6	7.33	10.8	12.18	25.8	39.8	0.0
Chile	2006	4.27		6.57		11.82		38.7	
	2012	5.98	40.0	7.41	12.7	13.26	12.2	46.1	19.2
	2018	6.22	4.0	8.96	21.0	13.95	5.2	47.3	2.7
Colombia	2006	5.25		5.91		10.53		41.9	
	2012	3.91	-25.4	5.91	0.0	11.24	6.7	40.4	-3.6
	2018	4.82	23.2	6.13	3.8	11.22	-0.1	41.9	3.7
México	2006	4.10		6.46		10.94		42.9	
	2012	4.96	20.9	6.56	1.5	10.72	-2.1	38.6	-9.9
	2018	4.27	-13.8	6.61	0.8	11.74	9.5	40.4	4.7
Uruguay	2006	7.55		6.89		12.40		43.9	
	2012	7.81	3.4	8.21	19.1	11.24	-9.3	38.6	-12.0
	2018	9.63	23.3	7.11	-13.3	11.85	5.4	39.6	2.5
<i>Reference country</i>									
Portugal	2006	12.31		8.38		9.59		41.3	
	2012	11.95	-3.0	9.68	15.6	11.01	14.7	42.5	2.8
	2018	11.41	-4.5	6.79	-29.9	12.69	15.3	48.0	13.0

would be encouraged to improve resource utilization. On the contrary, schools performing efficiently but achieving poor effectiveness, particularly in less developed economies, require help with resource allocation. Finally, schools with low effectiveness and low efficiency scores will need a combination of policies. The results reported below summarize the overall situation of schools at the aggregate (country) level. We then make cross-country and cross-time comparisons using PISA data and the methodology presented in Sect. 2, using Portugal 2006 as the reference technology. We also state some insights for national educational policy design.

Based on the distances in Table 3 and applying Eqs. (6) and (13), we compute performance and productivity gaps. Tables 4 and 5 show the PGI and base-group CDMI, together with their decompositions for 2018, respectively. A similar analysis was conducted for PISA 2006 and 2012 (see the four Tables included in “Appendix”).

The above tables indicate how each country on the vertical axis (Group B in the notation followed in Sect. 2) compares with each country on the horizontal axis (Group A in the notation followed in Sect. 2). For example, the PGI in Table 4

Table 3 DEA effectiveness and efficiency scores (averages)

Country	Wave	Country reference		Portugal 2006 reference technology (R)		Effectiveness/efficiency correlation
		Effectiveness	Efficiency	Effectiveness	Efficiency	
Argentina	2006	0.721	0.830	0.652	0.804	0.307
	2012	0.735	0.820	0.681	0.771	0.309
	2018	0.715	0.846	0.682	0.926	0.211
Brazil	2006	0.635	0.681	0.651	1.552	0.013
	2012	0.690	0.655	0.652	1.341	0.023
	2018	0.657	0.693	0.675	1.346	-0.097
Chile	2006	0.743	0.842	0.717	1.397	0.324
	2012	0.761	0.822	0.760	1.197	-0.186
	2018	0.809	0.838	0.767	1.053	0.107
Colombia	2006	0.746	0.805	0.660	1.335	0.299
	2012	0.704	0.789	0.687	1.555	0.097
	2018	0.746	0.724	0.698	1.431	0.047
Mexico	2006	0.748	0.728	0.720	1.542	0.015
	2012	0.752	0.641	0.747	1.486	-0.032
	2018	0.751	0.746	0.702	1.618	0.057
Uruguay	2006	0.763	0.807	0.737	1.091	0.222
	2012	0.730	0.852	0.693	0.942	-0.129
	2018	0.761	0.780	0.700	1.059	-0.024
<i>Reference</i>						
Portugal	2006	0.799	0.851	0.799	0.851	0.168
	2012	0.823	0.805	0.821	0.852	0.219
	2018	0.836	0.827	0.813	1.089	0.045

indicates that the outcome gap is 10% higher in Chile (CHL) than in Colombia (COL = 1.10) and 13.7% higher than in Brazil (BRA = 1/0.879).¹³ In the other two panels of Table 4, we learn that the gap between Chilean and Colombian school performance is the product of an effectiveness gap—Chilean schools were 8.4% more effective (EfG = 1.084), and had an outcome possibility set gap 1.4% better (OPSG = 1.014) than that their Colombian counterparts. Regarding the PGI with respect to Brazilian schools this is fully explained by the effectiveness gap 23.3% better in Chile because the outcome possibility set gap is dominated by Brazilian schools by 8.4%.

¹³ It is important to highlight that from now on all performance and productivity gaps and gaps changes results are obtained with the distances of Table 3 using equations from 5 to 16. For example, from Table 4 PGI we know that the performance gap between, let say, Uruguay and Argentina is 1.026 for 2018. This is the result of using Eq. (12). We divide the average effectiveness of Uruguayan schools with respect to Portugal 2006, according to Table 3 this value is 0.700, by the same distance for Argentina, 0.682. As for presentation purposes all values were rounded to three decimals, direct operations could bring about slight differences in the third decimal.

Therefore, we can use Tables 4 and 5 to compare any of the six Latin American countries. Table 4 shows that Chilean schools outperformed all other school samples in terms of performance in 2018. Average effectiveness (EfG) was higher than in all the other countries. The pattern is different for Brazil that combines the worst performance with outperforming all the other countries in terms of outcome possibility set gap (OPSG).

From Table 5, we learn that Argentinean schools were the least productive during 2018. This result is explained by a high technology gap with respect to other countries, which is not offset by the general positive technical efficiency gap in this year. This result is similar for Chile secondary schools, which are less efficient than schools from the other countries, except for Argentinean schools. Chilean technology is outperformed by the other countries' technologies, except for Argentina, where the technology gap favors Chilean educational production by 14.9% ($TG=1/0.871$).

Figure 1 illustrates the base-group CDMI and the base-group PGI for the analyzed countries and dimensions summarizing the major gaps found. This diagram shows the results reported in Table 3 with Portugal 2006 as the baseline.

In Fig. 1, performance gaps with respect to the base group reference are measured on the y-axis, whereas productivity gaps are captured on the x-axis. Within this framework, Portugal 2006 technology acts as the reference, occupying the origin

Table 4 Decomposition of base-group PGI* performance scores (PISA 2018)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PGI (Performance Gap index)</i>						
ARG	1.000	1.010	0.889	0.977	0.971	0.974
BRA	0.990	1.000	0.879	0.967	0.961	0.964
CHL	1.125	1.137	1.000	1.100	1.093	1.096
COL	1.023	1.034	0.909	1.000	0.994	0.997
MEX	1.030	1.041	0.915	1.006	1.000	1.004
URY	1.026	1.037	0.912	1.003	0.997	1.000
<i>EfG (effectiveness gap)</i>						
ARG	1.000	1.089	0.883	0.958	0.952	0.939
BRA	0.918	1.000	0.811	0.880	0.874	0.862
CHL	1.132	1.233	1.000	1.084	1.078	1.063
COL	1.044	1.137	0.922	1.000	0.994	0.980
MEX	1.050	1.144	0.928	1.006	1.000	0.986
URY	1.065	1.160	0.941	1.020	1.014	1.000
<i>OPSG (outcome possibility set gap)</i>						
ARG	1.000	0.928	1.006	1.020	1.020	1.037
BRA	1.078	1.000	1.084	1.099	1.099	1.118
CHL	0.994	0.923	1.000	1.014	1.014	1.031
COL	0.980	0.910	0.986	1.000	1.000	1.017
MEX	0.981	0.910	0.986	1.000	1.000	1.017
URY	0.964	0.894	0.969	0.983	0.983	1.000

*Portugal 2006 is the reference technology

Table 5 Decomposition of base-group CDMI* productivity scores (PISA 2018)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>CDMI (Camanho Dyson Malmquist index)</i>						
ARG	1.000	0.688	0.879	0.647	0.572	0.874
BRA	1.454	1.000	1.278	0.941	0.832	1.271
CHL	1.138	0.782	1.000	0.736	0.651	0.995
COL	1.546	1.063	1.359	1.000	0.884	1.351
MEX	1.748	1.202	1.536	1.131	1.000	1.528
URY	1.144	0.787	1.005	0.740	0.654	1.000
<i>EG (efficiency gap)</i>						
ARG	1.000	1.220	1.010	1.169	1.134	1.086
BRA	0.819	1.000	0.827	0.958	0.929	0.890
CHL	0.990	1.209	1.000	1.157	1.123	1.075
COL	0.856	1.044	0.864	1.000	0.970	0.929
MEX	0.882	1.076	0.891	1.031	1.000	0.958
URY	0.921	1.124	0.930	1.076	1.044	1.000
<i>TG (technology gap)</i>						
ARG	1.000	0.564	0.871	0.554	0.505	0.805
BRA	1.775	1.000	1.545	0.982	0.895	1.429
CHL	1.149	0.647	1.000	0.636	0.580	0.925
COL	1.806	1.018	1.572	1.000	0.911	1.455
MEX	1.982	1.117	1.725	1.097	1.000	1.596
URY	1.242	0.700	1.081	0.687	0.627	1.000

*Portugal 2006 is the reference technology

of the coordinates (1, 1) in Fig. 1. The reference technology can be assumed to be the production frontier for Portugal 2006.¹⁴ The graph has four quadrants, where countries in the north-east quadrant perform better in terms of effectiveness and efficiency than the reference, whereas countries in the South-West quadrant perform worse in both dimensions, countries in the south-east quadrant dominate the reference in terms of efficiency but not in terms of effectiveness, and countries in the north-west quadrant are less efficient but more effective than the reference. Most countries, except Argentina and Portugal in both periods and Uruguay 2012, fall into the south-east quadrant, showing a positive efficiency gap with respect to Portugal 2006 technology, but being less effective than this reference. In fact, we did not find any country dominating Portugal 2006 in terms of effectiveness.

Every country appears three times in the plot in order to represent the distance in effectiveness and efficiency in the three periods with respect to Portugal 2006 technology. At the same time, Fig. 1 compares the gap evolution of a country in three dimensions: against itself, against the analyzed set of Latin American

¹⁴ Note that under this approach it is also possible to compare the mean behavior of Portuguese schools in 2006 against their own technology. Obviously, in this case for Portugal 2006 all the distance with respect to the technology is explained by the efficiency component.

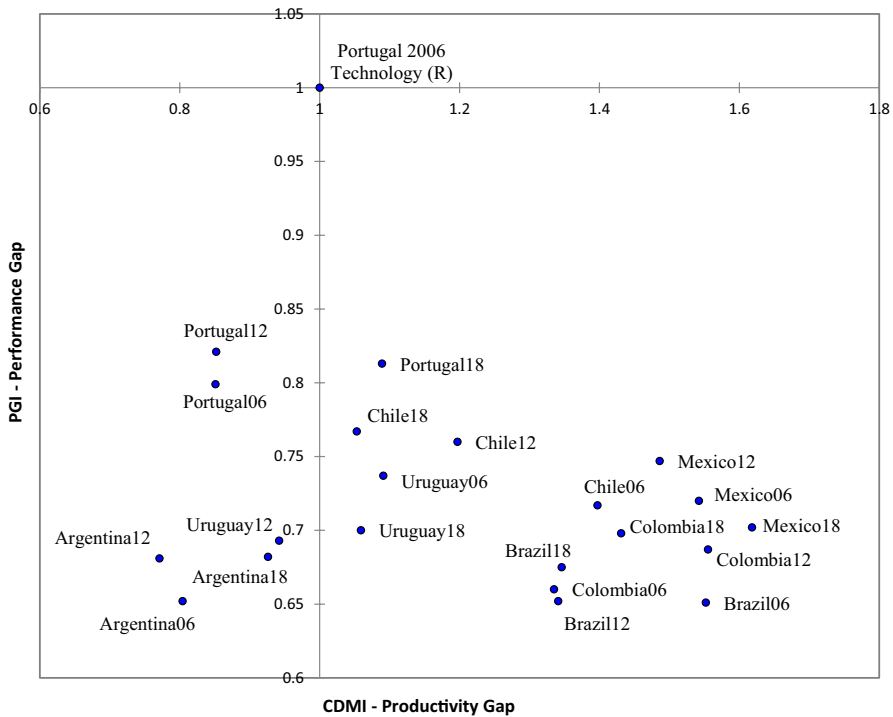


Fig. 1 PGI and CDMI. Performance and productivity gap evolution across countries (PISA 2006, 2012, 2018)

countries and against the reference. First, it is interesting to evaluate the development of a country with respect to itself from 2006 to 2018. Again, the most favorable direction is north-east because this result suggests a general improvement in effectiveness and efficiency. This is the direction followed by Colombia, Argentina and Portugal. With the same reasoning, the worst direction is south-west, where both dimensions deteriorate, as is the case for Uruguay. Second, a country that crosses an axis can be said to overtake (from west to east) or be overtaken (from east to west) by, the reference depending on the trajectory. Again, Uruguay's results confirm that the country was outperformed by the reference in 2012 with respect to 2006. Third, the distance, measured as the ratio between two countries in the same year, indicates the base-group PGI and the base-group CDMI that correspond with the performance and the productivity gaps. According to Eqs. (6) and (13), these gaps can be decomposed for explaining the drivers of these gaps that are discussed in the next section. Finally, we have also plotted the behavior for Portugal 2012 and 2018 in order to analyze the path followed by the reference. It is clear that effectiveness increased on average for Portugal during 2012 with respect to 2006. From 2012 to 2018 Portugal gained in productivity

Table 6 Pseudo-panel Malmquist index for performance (PGIC) and its decomposition into EfGC and OPSGC (PISA, 2006–2012)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PGIC (Performance Gap Index Change)</i>						
ARG	1.044	1.043	0.985	1.002	1.007	1.110
BRA	0.959	1.001	0.944	0.961	0.965	1.064
CHL	1.016	1.060	1.060	1.017	1.022	1.127
COL	0.998	1.040	0.983	1.042	1.004	1.107
MEX	0.993	1.037	0.978	0.996	1.037	1.103
URY	0.901	0.939	0.887	0.903	0.906	0.940
<i>EfGC (effectiveness gap change)</i>						
ARG	1.019	0.938	0.995	1.080	1.015	1.066
BRA	1.066	1.086	1.060	1.152	1.081	1.134
CHL	1.005	0.944	1.025	1.086	1.020	1.071
COL	0.926	0.868	0.920	0.944	0.940	0.986
MEX	0.986	0.925	0.980	1.064	1.005	1.050
URY	0.938	0.881	0.934	1.015	0.952	0.957
<i>OPSGC (outcome possibility set gap change)</i>						
ARG	1.024	1.112	0.990	0.928	0.993	1.043
BRA	0.899	0.922	0.892	0.835	0.893	0.938
CHL	1.010	1.122	1.034	0.937	1.001	1.053
COL	1.078	1.197	1.067	1.104	1.070	1.123
MEX	1.008	1.120	0.999	0.935	1.032	1.051
URY	0.959	1.066	0.950	0.890	0.951	0.982

at the cost of a slight declining in performance. However, Portugal continues being a valid benchmark for Latin American countries in terms of performance.

5 Productivity gaps over time

Finally, we report the PGIC and PPMI index computations, used to compare changes in performance and productivity for all countries across the three PISA waves, together with their respective decompositions applying the methodology described in Sects. 2.2 and 2.3. The results are reported in Tables 6, 7, 8 and 9 and illustrated in Figs. 2, 3 and 4. All the information contained in Tables 6, 7, 8 and 9 is summarized in Figs. 2, 3 and 4. We make a distinction between PGIC, PPMI (Fig. 2) and their respective decompositions into EfGC and EGC (Fig. 3) and into OPSGC and TGC (Fig. 4), plotting each country in order to represent the changes over the entire period 2006 to 2018.

Table 7 Pseudo-panel Malmquist index for performance (PGIC) and its decomposition into EfGC and OPSGC. (PISA, 2012–2018)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PGIC (Performance Gap Index Change)</i>						
ARG	1.002	0.967	0.992	0.987	1.066	0.992
BRA	1.034	1.035	1.026	1.020	1.102	1.025
CHL	1.008	0.975	1.009	0.995	1.074	0.999
COL	1.013	0.980	1.005	1.015	1.080	1.005
MEX	0.938	0.908	0.931	0.926	0.940	0.930
URY	1.008	0.976	1.001	0.995	1.075	1.010
<i>EfGC (effectiveness gap change)</i>						
ARG	0.973	1.023	0.915	0.918	0.974	0.933
BRA	0.978	0.951	0.895	0.898	0.952	0.912
CHL	1.092	1.117	1.063	1.003	1.064	1.019
COL	1.089	1.114	0.997	1.060	1.061	1.016
MEX	1.027	1.050	0.940	0.943	0.999	0.958
URY	1.072	1.096	0.981	0.984	1.044	1.043
<i>OPSGC (outcome possibility set gap change)</i>						
ARG	1.029	0.946	1.084	1.075	1.094	1.063
BRA	1.057	1.088	1.146	1.136	1.157	1.124
CHL	0.923	0.873	0.949	0.991	1.009	0.981
COL	0.930	0.880	1.009	0.958	1.018	0.989
MEX	0.914	0.864	0.991	0.982	0.941	0.971
URY	0.941	0.890	1.020	1.011	1.030	0.968

Each country is plotted¹⁵ in Figs. 2, 3 and 4 according to (11) and (16), whereas the distances between two countries regarding the two dimensions/axes considered represent the gap change measured in (9) and (14). There is no clear pattern followed by the six countries in Fig. 2. We observe how Colombia and Argentina are on the north-east quadrant showing that in these two countries performance and productivity improved during this twelve-years period. The reverse is Uruguay (south-west), the only country where both indicators decrease over the 2006 to 2018 period. Departing from an advantageous position in 2006, Uruguayan schools have lost their relative better situation with respect to other countries. Like Uruguay, other countries (north-west) also experienced a productivity loss. This is the case of Chile and Brazil, however, their performance with respect to 2006 increased being Chile the country with the highest change. In the south-east quadrant is Mexico, indicating that this country gained in productivity at the cost of losing performance. Summarizing, the current poor results in 2006 single out the effectiveness dimension as

¹⁵ Note that applying the circularity property of Eqs. (10), (11), (15) and (16), the exact coordinates for dots plotted in Figs. 2, 3 and 4 showing all performance and productivity changes and their decompositions between 2006 and 2018 are obtained multiplying the main diagonal in Tables 6 and 8 (2006–2012 changes) by the main diagonal in Tables 7 and 9 (2012–2018 changes) respectively. The values outside the main diagonals in Tables from 6, 7, 8 and 9 represent the distances between them.

Table 8 Pseudo-panel Malmquist index (PPMI) and its decomposition into EGC and TGC. (PISA. 2006–2012)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PPMI (Pseudo-panel Malmquist index for productivity)</i>						
ARG	0.959	1.110	1.120	0.824	0.996	1.111
BRA	0.901	0.864	1.008	0.743	0.897	1.001
CHL	0.893	0.992	0.857	0.736	0.889	0.992
COL	1.214	1.346	1.358	1.164	1.208	1.349
MEX	1.004	1.115	1.125	0.828	0.964	1.117
URY	0.900	0.999	1.008	0.741	0.895	0.863
<i>EGC (efficiency gap change)</i>						
ARG	0.988	1.028	1.013	1.009	1.123	0.935
BRA	0.973	0.961	0.985	0.982	1.092	0.909
CHL	0.987	1.015	0.976	0.996	1.109	0.923
COL	0.991	1.018	1.004	0.979	1.113	0.927
MEX	0.891	0.916	0.902	0.898	0.880	0.834
URY	1.070	1.100	1.083	1.079	1.199	1.057
<i>TGC (technology gap change)</i>						
ARG	0.970	1.080	1.104	0.817	0.886	1.189
BRA	0.926	0.899	1.023	0.756	0.820	1.100
CHL	0.905	0.977	0.878	0.739	0.802	1.075
COL	1.224	1.323	1.353	1.189	1.086	1.456
MEX	1.128	1.219	1.247	0.921	1.095	1.341
URY	0.841	0.909	0.930	0.687	0.746	0.817

the priority, and this causes more inputs to be set aside to improve test scores. Four countries achieved better performance but for two of them (Chile and Brazil) this was in detriment to the productivity dimension.

Figures 3 and 4 show the PGIC and PPMI decompositions, which identifies the countries that benefit most from effectiveness and efficiency changes (EfGC vs. EGC), together with the outcome possibility set gap change and technological gap change (OPSGC vs. TGC). Figure 4 shows how the good performance and productivity results obtained by Colombia and Argentina (Fig. 2) are driven by shifting up their production frontiers. Their results in terms of EfGC and EGC are not so positive but this appears to be a sound result because when some schools shift up the frontier, the remaining schools need some time to catch up with respect to the best performers. In Fig. 3 we can understand that the high positive performance gap change in Chile (Fig. 2) is driven by a high effectiveness gap change, i.e. in 2018 all Chilean schools are closer to their outcome possibility set than in 2006 even although the outcome possibility set gap is slightly backwards compared to 2006 (Fig. 4). This seems to be an alternative way of increasing performance at the cost of a loss in productivity. A possible explanation, as we confirm seeing descriptive statistics of Chile in Table 2, is that expending relatively more resources than in other countries leads to a loss in both productivity components, efficiency and technology

Table 9 Pseudo-panel Malmquist index (PPMI) and its decomposition into EGC and TGC. (PISA. 2012–2018)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PPMI (Pseudo-panel Malmquist index for productivity)</i>						
ARG	1.201	1.196	1.365	1.305	1.103	1.068
BRA	0.836	1.004	1.141	1.091	0.922	0.893
CHL	0.732	0.876	0.880	0.956	0.808	0.782
COL	0.766	0.917	1.046	0.921	0.846	0.819
MEX	0.906	1.084	1.237	1.183	1.089	0.968
URY	0.936	1.120	1.278	1.222	1.033	1.125
<i>EGC (efficiency gap change)</i>						
ARG	1.032	0.975	1.012	1.123	0.886	1.128
BRA	1.026	1.059	1.038	1.153	0.909	1.158
CHL	0.989	0.963	1.020	1.111	0.875	1.115
COL	0.890	0.868	0.900	0.918	0.788	1.004
MEX	1.129	1.100	1.142	1.268	1.165	1.274
URY	0.886	0.864	0.897	0.996	0.785	0.915
<i>TGC (technology gap change)</i>						
ARG	1.164	1.228	1.350	1.162	1.246	0.947
BRA	0.814	0.948	1.099	0.946	1.015	0.771
CHL	0.741	0.910	0.863	0.861	0.923	0.702
COL	0.861	1.057	1.162	1.002	1.073	0.815
MEX	0.803	0.985	1.083	0.932	0.935	0.760
URY	1.056	1.297	1.425	1.227	1.316	1.230

but allows to boot performance. Finally, we should highlight that a simultaneous positive EfGC and EGC, as in Brazil and Mexico (Fig. 3), do not imply higher performance and productivity.

6 Conclusions

The main objective of this paper is to propose a tool for measuring the performance and productivity gaps across a set of production units belonging to different groups for monitoring their evolution and analyzing their components. To do this, for measuring productivity gaps we use the approach proposed by Aparicio and Santín (2018), which is grounded on a base-group base-period index. Additionally, for the first time, we introduce an index to measure performance gaps and its decomposition. As an empirical illustration of the approach, we focus our attention on the educational sector. In particular, we analyzed a set of six Latin American countries over time (2006 and 2018).

From the analysis, we find that the simultaneous analysis of performance and productivity gaps together with their evolution over time is an accurate way of benchmarking countries for monitoring improvements and weaknesses in education

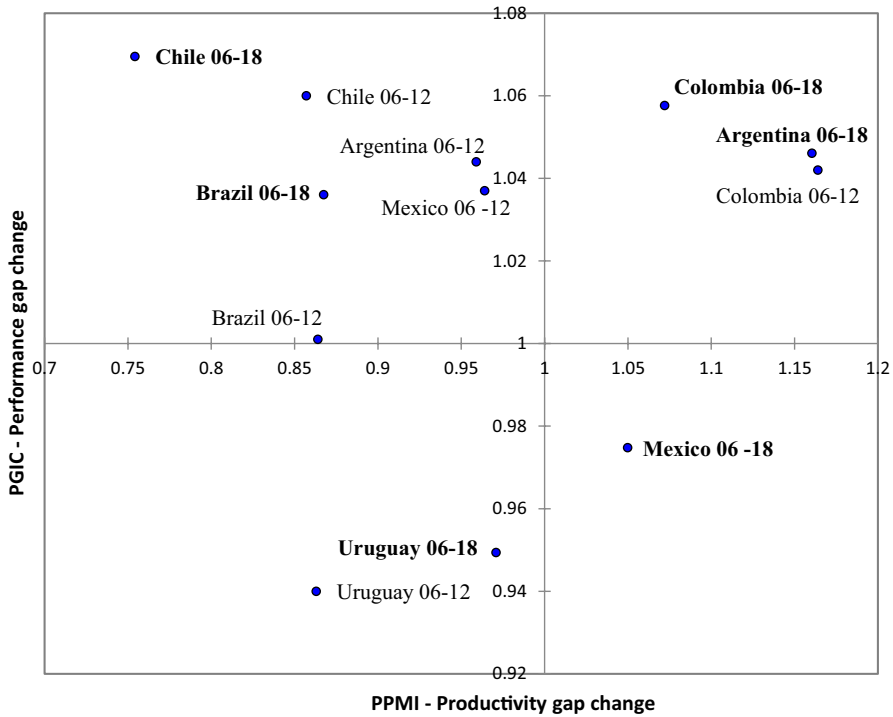


Fig. 2 PGIC and PPMI. Performance and productivity gap change (PISA, 2006–2012; PISA, 2006–2018)

systems. New policies, programs, resources and practices aimed at enhancing school performance and productivity should be monitored in comparison with other actors to detect, learn and replicate best practices. At the same time, performance and productivity are different dimensions that, in education at least, should be considered simultaneously. If we only evaluate taking one of the two dimensions, we might conclude that the educational system is working, leading to imprecise policy recommendations. For policymakers, it is necessary to understand the cost of performance or productivity losses. Countries with poor performance (in terms of test scores) should probably focus first on this dimension setting as priority boosting students' achievements. Later, once they have reached a specified threshold, the focus could switch to productivity. However, both aspects should be taken into account at all times.

We should underscore some of the results of our analysis. First, the pattern for most countries in Fig. 2 appears to show that this is what is occurring in most Latin American countries, where effectiveness is the priority and is being pushed upwards by the allocation of relatively more inputs to this aspect. In some cases, as in Colombia and Argentina, this is done in an efficient way while in others, as in Chile or Brazil, this is leading to cuts in the efficiency dimension. Second, performance gaps among Latin American countries are relatively low, although, in 2018, Chile is the benchmark for enhancing educational practices in the region.

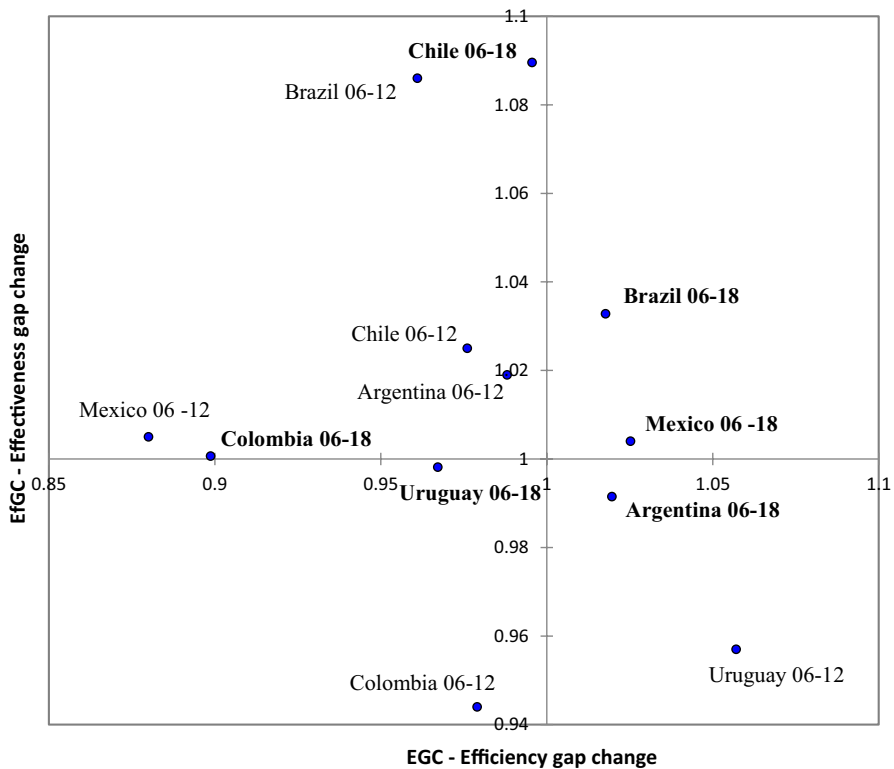


Fig. 3 EfGC and EGC. Effectiveness and efficiency gap change (PISA 2006–2012; PISA 2006–2018)

Uruguay, which was the benchmark in 2006, has been reached by other countries and is the only country where both dimensions have decreased over the period 2006–2018. Third, Mexico and Colombia clearly achieve the best results with respect to the efficiency dimension. This means that these countries are getting their results with relatively fewer resources than the other countries.

This approach can be used to address many research lines in the near future. As we have already highlighted, this tool could be used not only for the follow-up of schools in different countries but also for analyzing the evolution of other differences between groups of production units working within different environments or under different educational policies. The information gathered could be used by policy makers to justify new programs or new framework laws to boost results in either or both dimensions. This is especially important in the case of countries with federal systems of government to assure that educational inequalities among states are tolerable or for making decisions on investment in public or stated-funded private schools.

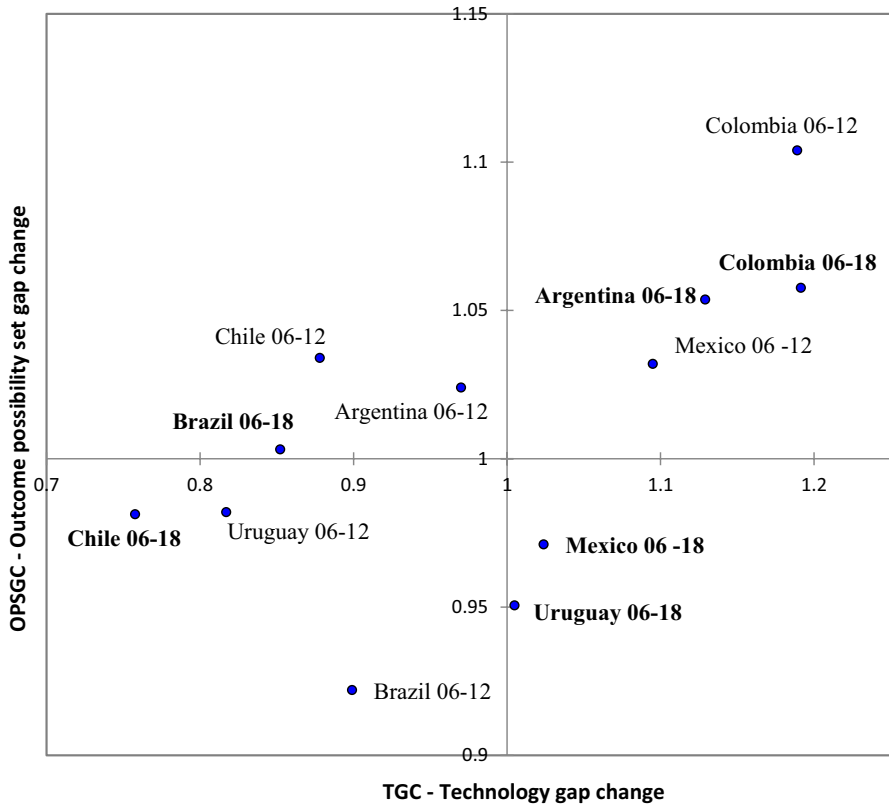


Fig. 4 OPSGC and TGC. Outcome possibility set gap change and technology gap change (PISA, 2006–2012; PISA 2006–2018)

Acknowledgements We thank two anonymous reviewers for providing constructive comments and help in improving the contents and presentation of this paper. Additionally, J. Aparicio thanks the Spanish Ministry for Economy and Competitiveness (*Ministerio de Economía, Industria y Competitividad*), the State Research Agency (*Agencia Estatal de Investigación*) and the European Regional Development Fund (*Fondo Europeo de Desarrollo Regional*) for financial support under grant PID2019-105952GB-I00 (AEI/FEDER, UE). D. Santín acknowledges funding received from the Spanish Ministry for Economy and Competitiveness (*Ministerio de Economía, Industria y Competitividad*) project referenced ECO2017-83759-P.

Compliance with ethical standards

Conflict of interest The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Appendix

See Tables 10, 11, 12 and 13.

Table 10 Decomposition of PGI effectiveness scores (PISA 2006)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PGI (Performance Gap index)</i>						
ARG	1	1.001	0.909	0.988	0.905	0.885
BRA	0.999	1	0.908	0.986	0.904	0.884
CHL	1.100	1.101	1	1.087	0.995	0.973
COL	1.012	1.014	0.920	1	0.916	0.896
MEX	1.105	1.106	1.005	1.092	1	0.978
URY	1.130	1.131	1.028	1.116	1.022	1
<i>EfG (effectiveness gap)</i>						
ARG	1	1.135	0.97	0.966	0.963	0.945
BRA	0.881	1	0.855	0.851	0.849	0.833
CHL	1.031	1.170	1	0.995	0.993	0.974
COL	1.035	1.175	1.005	1	0.997	0.979
MEX	1.038	1.178	1.007	1.003	1	0.981
URY	1.058	1.200	1.027	1.021	1.019	1
<i>OPSG (outcome possibility set gap)</i>						
ARG	1	0.882	0.937	1.023	0.939	0.936
BRA	1.134	1	1.061	1.159	1.064	1.061
CHL	1.067	0.943	1	1.092	1.003	0.999
COL	0.978	0.863	0.916	1	0.918	0.915
MEX	1.065	0.940	0.997	1.089	1	0.996
URY	1.068	0.943	1.001	1.093	1.004	1

Table 11 Decomposition of CDMI efficiency scores (PISA 2006)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>CDMI (Camanho Dyson Malmquist index)</i>						
ARG	1	0.518	0.575	0.602	0.521	0.737
BRA	1.931	1	1.111	1.162	1.006	1.422
CHL	1.739	0.900	1	1.046	0.906	1.281
COL	1.661	0.861	0.956	1	0.866	1.224
MEX	1.919	0.994	1.104	1.155	1	1.413
URY	1.357	0.703	0.781	0.817	0.708	1
<i>EG (technical efficiency gap)</i>						
ARG	1	1.218	0.985	1.031	1.14	1.029
BRA	0.821	1	0.809	0.846	0.936	0.845
CHL	1.015	1.236	1	1.046	1.157	1.044
COL	0.970	1.182	0.956	1	1.106	0.998
MEX	0.877	1.068	0.864	0.904	1	0.902
URY	0.972	1.183	0.958	1.002	1.109	1
<i>TG (technology gap)</i>						
ARG	1	0.425	0.584	0.584	0.457	0.716
BRA	2.353	1	1.373	1.373	1.075	1.684
CHL	1.712	0.728	1	1	0.783	1.227
COL	1.712	0.728	1.000	1	0.783	1.226
MEX	2.188	0.930	1.277	1.277	1	1.566
URY	1.397	0.594	0.815	0.816	0.639	1

Table 12 Decomposition of base-group PGI* effectiveness scores (PISA 2012)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>PGI (Performance Gap index)</i>						
ARG	1	1.044	0.895	0.990	0.911	0.982
BRA	0.958	1	0.857	0.948	0.872	0.941
CHL	1.117	1.167	1	1.106	1.017	1.097
COL	1.010	1.055	0.904	1	0.920	0.992
MEX	1.098	1.147	0.983	1.087	1	1.079
URY	1.018	1.063	0.912	1.008	0.927	1
<i>EfG (effectiveness gap)</i>						
ARG	1	1.065	0.965	1.043	0.977	1.007
BRA	0.939	1	0.906	0.980	0.918	0.945
CHL	1.036	1.104	1	1.081	1.013	1.043
COL	0.959	1.020	0.925	1	0.937	0.965
MEX	1.024	1.089	0.987	1.067	1	1.030
URY	0.993	1.058	0.959	1.036	0.971	1
<i>OPSG (outcome possibility set gap)</i>						
ARG	1	0.981	0.928	0.949	0.932	0.976
BRA	1.019	1	0.946	0.968	0.950	0.995
CHL	1.078	1.057	1	1.023	1.004	1.052
COL	1.054	1.033	0.978	1	0.982	1.028
MEX	1.073	1.053	0.996	1.018	1	1.047
URY	1.025	1.005	0.951	0.973	0.955	1

*Portugal 2006 is the reference technology

Table 13 Decomposition of base-group CDMI* efficiency scores (PISA 2012)

Country	ARG	BRA	CHL	COL	MEX	URY
<i>CDMI (Camanho Dyson Malmquist index)</i>						
ARG	1	0.575	0.644	0.496	0.519	0.819
BRA	1.739	1	1.120	0.863	0.902	1.424
CHL	1.553	0.893	1	0.770	0.805	1.271
COL	2.016	1.159	1.299	1	1.046	1.651
MEX	1.927	1.109	1.242	0.956	1	1.579
URY	1.221	0.702	0.787	0.606	0.633	1
<i>EG (technical efficiency gap)</i>						
ARG	1	1.252	0.998	1.040	1.280	0.962
BRA	0.799	1	0.797	0.831	1.022	0.768
CHL	1.002	1.255	1	1.042	1.283	0.964
COL	0.962	1.203	0.960	1	1.231	0.925
MEX	0.781	0.978	0.779	0.812	1	0.752
URY	1.040	1.302	1.037	1.081	1.330	1
<i>TG (technology gap)</i>						
ARG	1	0.459	0.645	0.477	0.405	0.851
BRA	2.179	1	1.405	1.038	0.882	1.853
CHL	1.550	0.712	1	0.739	0.628	1.319
COL	2.096	0.963	1.353	1	0.850	1.785
MEX	2.469	1.134	1.592	1.176	1	2.100
URY	1.175	0.540	0.758	0.560	0.476	1

*Portugal 2006 is the reference technology

References

- Afonso A, St Aubyn M (2006) Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs. *Econ Model* 23(3):476–491
- Amado CA, Santos SP, São José JM (2018) Measuring and decomposing the gender pay gap: a new frontier approach. *Eur J Oper Res* 271(1):357–373
- Amado CAF, Barreira AP, Santos SP, Guimarães MH (2019) Comparing the quality of life of cities that gained and lost population: an assessment with DEA and the Malmquist index. *Pap Region Sci* 98(5):2075–2097
- Aparicio J, Santín D (2018) A note on measuring group performance over time with pseudo-panels. *Eur J Oper Res* 267(1):227–235
- Aparicio J, Crespo-Cebada E, Pedraja-Chaparro F, Santín D (2017) Comparing school ownership performance using a pseudo-panel database: a Malmquist–type index approach. *Eur J Oper Res* 256:533–542
- Aristovnik A (2012) The relative efficiency of education and R&D expenditures in the new EU member states. *J Bus Econ Manag* 13(5):832–848
- Aristovnik A, Obadić A (2014) Measuring relative efficiency of secondary education in selected EU and OECD countries: the case of Slovenia and Croatia. *Technol Econ Dev Econ* 20(3):419–433
- Berg SA, Forsund FR, Jansen ES (1992) Malmquist indices of productivity growth during the deregulation of Norwegian Banking, 1980–89. *Scand J Econ* 94((Supplement)):211–228
- Camanho AS, Dyson RG (2006) Data envelopment analysis and Malmquist indices for measuring group performance. *J Prod Anal* 26:35–49

- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2:429–444
- Cherchye L, Moesen W, Puyenbroeck T (2004) Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. *JCMS J Common Mark Stud* 42(5):919–955
- Cherchye L, De Witte K, Perelman S (2019) A unified productivity-performance approach applied to secondary schools. *J Oper Res Soc* 70(9):1522–1537
- Cordero JM, Polo C, Santín D, Simancas R (2018) Efficiency measurement and cross-country differences among schools: a robust conditional nonparametric analysis. *Econ Model* 74:45–60
- De Jorge J, Santín D (2010) Determinantes de la eficiencia educativa en la Unión Europea. *Hacienda Pública Española* 193:131–155
- De Witte K, López-Torres L (2017) Efficiency in education: a review of literature and a way forward. *J Oper Res Soc* 68(4):339–363
- Deutsch J, Dumas A, Silber J (2013) Estimating an educational production function for five countries of Latin America on the basis of the PISA data. *Econ Educ Rev* 36:245–262
- Farrell MJ (1957) The measurement of productive efficiency. *J R Stat Soc Ser A (General)* 120(3):253–290
- Ferreira D, Marques RC (2015) Did the corporatization of Portuguese hospitals significantly change their productivity? *Eur J Health Econ* 16(3):289–303
- Ganzeboom H, De Graaf P, Treiman J, De Leeuw J (1992) A standard international socio-economic index of occupational status. *Soc Sci Res* 21(1):1–56
- Hanushek EA, Kimko DD (2000) Schooling, labor-force quality, and the growth of nations. *Am Econ Rev* 90(5):1184–1208
- Hanushek EA, Woessmann L (2008) The role of cognitive skills in economic development. *J Econ Liter* 46(3):607–668
- Hanushek EA, Woessmann L (2011a) The economics of international differences in educational achievement. *Handbook of the Economics of Education*, vol 3. Elsevier, New York, pp 89–200
- Hanushek EA, Woessmann L (2011b) How much do educational outcomes matter in OECD countries? *Econ Policy* 26(67):427–491
- Hanushek EA, Woessmann L (2012) Schooling, educational achievement, and the Latin American growth puzzle. *J Dev Econ* 99(2):497–512
- Hanushek EA, Woessmann L (2020) Education, knowledge capital, and economic growth. *The Economics of Education*, 2nd edn. Academic Press, New York, pp 171–182
- Hanushek EA, Link S, Woessmann L (2013) Does school autonomy make sense everywhere? Panel estimates from PISA. *J Dev Econ* 104:212–232
- Kaarsen N (2014) Cross-country differences in the quality of schooling. *J Dev Econ* 107:215–224
- Lovell CAK (2003) The decomposition of Malmquist productivity indexes. *J Prod Anal* 20:437–458
- Lovell CAK, Pastor JT (1999) Radial DEA models without inputs or without outputs. *Eur J Oper Res* 118:46–51
- Mbuvu D, De Witte K, Perelman S (2012) Urban water sector performance in Africa: a stepwise bias-corrected efficiency and effectiveness analysis. *Util Policy* 22:31–40
- Melyn W, Moesen W (1991) Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. *Public Economics Research Paper No. 17*, Katholieke Universiteit Leuven, Leuven, Belgium
- OECD (1999) *Classifying educational programmes, Manual for ISCED-97 implementation in OECD countries*. OECD, Paris
- OECD (2005) *Handbook on constructing composite indicators*. OECD, Paris
- OECD (2013) *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. PISA. OECD Publishing, Paris
- OECD (2014a). *PISA 2012 results: What students know and can do. Student performance in mathematics, reading and science: Vol I*. OECD Publishing, revised edition, PISA
- OECD (2014b). *PISA 2012 Technical Report*. OECD Publishing
- Prieto AM, Zofio JL (2001) Evaluating effectiveness in public provision of infrastructure and equipment: the case of Spanish municipalities. *J Prod Anal* 15(1):41–58
- Santín D, Sicilia G (2015) Measuring the efficiency of public schools in Uruguay: main drivers and policy implications. *Latin Am Econ Rev* 24:5. <https://doi.org/10.1007/s40503-015-0019-5>
- Thanassoulis E, Shiraz RK, Maniadakis N (2015) A cost Malmquist productivity index capturing group performance. *Eur J Oper Res* 241(3):796–805
- Thieme C, Giménez V, Prior D (2012) A comparative analysis of the efficiency of national education systems. *Asia Pac Educ Rev* 13(1):1–15

Vaz CB, Camanho AS (2012) Performance comparison of retailing stores using a Malmquist-type index. *J Oper Res Soc* 63(5):631–645

World Bank (2020) World Development Indicators. (<https://databank.worldbank.org/source/world-development-indicators>)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.