Check for
updates

A Composite Indicator to Compare the Performance of Male and Female Students in Educational Systems

M. M. Segovia-González[1] · I. Contreras[1]

Abstract

Composite indicators have proved to be a suitable tool for the evaluation of complex phenomena, by enabling a large amount of information to be concentrated in a unique value. In this paper, a methodology based on geometric aggregation and free-selection of weights is proposed for the evaluation of the educational systems of OECD countries. In an effort to carry out a global evaluation, not only is a group of indicators considered to measure the academic outcomes, but also a group of indicators that measure the social dimension of the educational system and the self-perceptions of the students regarding their academic stage. The features of the proposed methodology enable the results for male and female students to be compared separately, since not only identify the contribution of each aspect to the global indicator, but also ascertain the sources of the differences between the values obtained for the male and female subgroups.

Keywords Education · Composite indicator · Data envelopment analysis · Gender perspective

1 Introduction

In multiple contexts, composite indicators have been shown to be valuable instruments for the representation of complex phenomena, by highlighting any existing inequalities, helping to generate attention and analysis, and enabling the progress to be monitored. A variety of entities have considered this instrument in their own way, in order to synthesise information and to ultimately generate internationally comparable measures.

A composite indicator should verify certain basic features. Firstly, it must include a solid foundation in theory that supports the aggregation of individual measures and the adequate representation of the target pursued by the composite indicator. The proposed indicator must also be operationally viable and easily replicable. Finally, the comparison

✉ M. M. Segovia-González
mmseggon@upo.es

I. Contreras
iconrub@upo.es

1   Department of Economics, Quantitative Methods and Economic History, Universidad Pablo de Olavide, Carr. de Utrera Km. 1, 41013 Seville, Spain

across the entities should be guaranteed, depending on the context, whether it be across regions, entities, etc.

The connection of education and composite indicators would require a wide review of the literature. An exhaustive literature review of this topic is beyond the purpose of this study. It is interesting to highlight that education performance has been considered traditionally integrated into composite indicators that propose the measure of development or inequalities in comparative studies across countries (e.g., Ferrara & Nisticò, 2013; Mitrovic et al., 2016; Monte & Schoier, 2022). But composite indicators have also been shown to be a valuable tool in the analysis of education performance. Hence, several studies have proposed comparative analysis of educational systems across countries (e.g., Stumbriene et al., 2020; Jovanovic-Gavrilovic & Radivojevic, 2017; Villar, 2017; Dominguez et al., 2021) or an analysis of particular aspects like higher education. In this line, most of the rankings existing nowadays have developed and used composite indicators to provide rankings of HE institutions or countries (among others, the Times Higher Education World University Ranking, the World University Ranking or the one developed by Torres-Salinas et al. (2011)). In a similar sense, the proposal of constructing composite indicators in order to analyse the quality of universities can be found in, among others Murias et al. (2007), El Gibari et al. (2018) (for the Spanish public university system), Benito and Romera (2011), Gnaldi and Ranalli (2016), Docampo and Cram (2017) and Bas and Carot (2020).

The proposal of comparable indicators with a gender perspective has received increasing attention in recent decades. The objective of a gender indicator involves the measurement of gender-related changes over time or across entities. According to Moser (2007), gender indicators can refer to quantitative indicators based on sex disaggregated statistical data, which provides separate measures for men and women, but can also capture qualitative changes. Measurements of gender equality might address changes in the relations between men and women, the outcomes of a particular policy, programme, or activity for women and men, or changes in the status or situation of men and women.

In the field of education, the inclusion of gender indicators plays a significant role. A number of indicators have been proposed to measure the literacy gap between men and women, and male and female enrolment rates in primary, secondary and tertiary education, among others. Traditionally, the consideration of gender perspective in education has focused on inequality of access for female students, the relevance of the curriculum in teaching gender perspective and the obstacles to the inclusion of such a perspective.

The interest in the inclusion of gender-indicators in education has largely been related with the study of development, how gender inequalities in education can hinder development, or, equivalently, how the promotion of equality in education can break the poverty cycle by reducing economic inequalities and promoting gender equality. In this regard, The Sustainable Development Goals (United Nations, 2020), include, as one of their targets, "to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all by 2030, and eliminate gender disparities and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations". In a similar aim, The Incheon Declaration recognises "the importance of gender equality in achieving the right to education for all... supporting gender-sensitive policies, planning and learning environments; mainstreaming gender issues in teacher training and curricula; and eliminating gender-based discrimination and violence in schools" (UNESCO, 2015). This report includes the target of gender equality for the 2030 Agenda in a central position. In most studies that propose the use of composite indicators, the inclusion of education and gender is justified either as a part of a complex index to measure inequalities (e.g., Bericat &

Sanchez-Bermejo, 2016; Ferrant, 2014; Bericat, 2012), to analyse the connection between education performance (e.g., Hadjar & Buchmann, 2016; Blossfeld et al., 2016; Arcos et al., 2007) or as a contextual aspect to make comparisons among the results (e.g., INES project from (OECD, 2021b)).

Defining gender inequality and gender equality in schooling entails more than a description of the numbers of girls and boys enrolled in and progressing through stages of instruction. The definition of inequality derived from Sen (1999) points to the existence of limits and constraints on opportunities.

In this paper we do not propose an analysis from a gender-perspective in the aforementioned sense, nor do we try to measure the gender-inequalities by computing the limitation in the access, the differences in the contents of the classes, or the discrimination in the classrooms. We strive to ascertain whether the evaluation of the outcomes from national educational systems presents differences when the evaluation is carried out on girls and boys separately, that is, whether the performance in academic results and in the social-transforming role of the systems differs for boys and girls. To this end, a methodology for the construction of composite indicators is proposed, and the evaluation of the educational systems for the OECD countries is carried out. By comparing the results of male and female students, considered as two separate subgroups of the same population, we strive to shed light on the matter and evaluate whether significant differences appear between genders, when the outcomes of education, as a broad concept, are evaluated.

Hence, the main target of the paper is to present a case study. To that end, we have considered an ad hoc methodology obtained as an adaptation of an existing procedure, that could be extended to other contexts which similar characteristics. With the proposed methodology, we ensure that all the objectives of the study are achieved. Firstly, to determine a composite indicator with an objective set of weights. Secondly, we propose a composite indicator developed under an aggregation scheme that places special emphasis on those aspects in which the unit under evaluation presents a poor indicator.

One unit with unbalanced values in the sub-indicators would opt for a compensatory aggregation scheme (like the additive aggregation) since such schemes allow a perfect substitutability among indicators. On the contrary, the geometric aggregation scheme inhibits this perfect transference of values from the highest indicators to the lowest ones. This characteristic leads the decision-makers to concentrate their effort on those measures that will have a greater impact on the improvement of the aspect most undervalued. It is important to bear in mind that the marginal increase of the global evaluation derived from an improvement in the lowest-valued indicators is larger than that which would be obtained from an increase in a better-valued indicator.

In addition, the geometric nature of the indicator will permit comparing the results for separate subgroups and identifying the main sources of comparative advantage from one unit to another or from one subgroup to the other for the same unit. This information about the performance of the units will be also helpful in the decision-making process. As will be discussed later, the nature of the proposed indicator permits insulating the sources of advantages (or disadvantages) of one unit with respect to others or from one subgroup with respect to the other. We can determine the individual effects generated from the comparison of the individual observed values, the weighting schemes and the benchmark considered as a basis on the aggregated value obtained for each unit when a comparative analysis is performed. In the light of these results, the decision-maker can detect if the weakness of the unit (or the group) is derived from their own observed values (in such cases, actions should be addressed to improve those values) or, in contrast, from the benchmarks or the

weighting schemes. In those cases, the effort should be directed to improving the complete set of sub-indicators.

The rest of the paper is organised as follows. Section 2 summarises the methodological procedure for the construction of the composite indicator. Section 3 incorporates the application of the proposed methodology into the evaluation of the educational systems in the OECD countries. Finally, Sect. 4 is devoted to the main concluding remarks and political implications.

2 A Geometric Composite Indicator Under the Benefit-of-the-Doubt Principle

From a methodological point of view, a composite indicator (CI) is a mathematical aggregation of a set of individual indicators (often referred to as subindicators), for the measurement of multidimensional concepts that cannot be captured by only one single indicator (OECD, 2008). The process of generating a composite indicator implies several successive decisions (OECD, 2008): selection of initial indicators, the way in which they are conceptually grouped (generating dimensions which cluster individual indicators), the decision on the data normalisation method, and the choice of the method to weight and aggregate subindicators. A complete revision of the concepts related with the construction of CIs can be seen, among others, in Nardo et al. (2005a), Nardo et al. (2005b).

The weighting and aggregation phases are two of the most determinants phases in this process (Esty et al., 2005; Saisana & Tarantola, 2002). In this section, a methodology based on a geometric aggregation scheme is proposed, in which the weighting factors are freely determined. The so-called *Benefit of the Doubt* (BoD) principle is taken as reference Cherchye et al. (2003), Cherchye et al. (2007), Murias et al. (2007) to determine the weighting scheme. As will be discussed later, Data Envelopment Analysis (DEA) is considered as a tool to determine weighting profiles for composite indicators.

In order to introduce the proposed methodology, let us consider a generic set of n units of alternatives that are to be evaluated by means of a CI. Let us also suppose that a set of m individual indicators have been collected. The individual indicator r with respect to the alternative i is denoted as I_{ri}, with $i = 1, 2, \ldots, n$ and $r = 1, \ldots, m$. The composite indicator for the alternative i is denoted as CI_i. For the sake of simplicity, we assume that all the individual indicators I_{ri} have been treated in such a way that the higher the value, the better the alternative.

In order to determine the CI_i values, a multiplicative aggregation scheme is proposed such that:

$$CI_i = \prod_{r=1}^{m} I_{ri}^{\omega_r}, \omega_r \geq 0, \sum_{r=1}^{m} \omega_r = 1, \tag{1}$$

where ω_r denotes the normalised weighting factor associated to the rth individual indicator.

Several authors have studied the advantages of geometric aggregation over the classical additive scheme. In Ebert and Welsch (2004) and Zhou et al. (2010), the authors point out several desirable properties, such as the scale-invariance and a lower degree of compensation between individual indicators. The latter supposes that this aggregation function penalises those alternatives with lower values in certain individual indicators (Yoon & Hwang, 1995). Consequently, a modification in an originally low-value indicator, would cause a greater variation in the CI than in high-value indicators. That is to say, the alternatives are

encouraged to improve their weaknesses rather than reinforce those aspects in which they are top performers. Other authors, see Zhou et al. (2006, 2010) among others, found that geometric aggregation implies a lower information loss.

The selection of the weighting scheme is inspired from the *Benefit of the doubt* (BoD) principle Cherchye et al. (2007), which is based on DEA methodology (Banker et al., 1984; Charnes et al., 1978). This family of procedures permits each alternative to select its own vector of weights. The underlying idea is to allow each alternative to maximise its composite value, under a set of common constraints in order to guarantee that all the values are limited. This free selection of weights enables each unit to use its own weight set Murias et al. (2007). Two major benefits are derived from these models. First, since the weighting values are adapted to unit measures of the sub-indicators, the application of a normalisation method is not necessary. Secondly, the CIs are computed with an objectively determined vector of weights and are not derived from a set of subjective decisions of the analyst. Several recent procedures, which combine both multiplicative aggregation and BoD, can be consulted in Blancas et al. (2012), Giambona and Vassallo (2014), Zhou et al. (2010), Van Puyenbroeck and Rogge (2017), Verbunt and Rogge (2018) and Dominguez et al. (2021), among others.

In this paper, a procedure based on the indirect CI-framework is developed as proposed in Van Puyenbroeck and Rogge (2017). The authors propose the construction of a CI with geometric aggregation using the weights derived from a DEA-inspired model. Hence, the procedure implies a number of successive steps. Firstly, the weighting scheme is determined using a BoD model. The sub-indicators are subsequently aggregated using these weighting factors in a multiplicative scheme.

As a first step, a max–min normalisation process is applied to guarantee the comparison between sub-indicators, thereby determining $I_{ir}^N \in [1, 2]$. To determine the weights associated to sub-indicator r, the revisited optimistic/pessimistic DEA model is considered, as developed in Contreras and Hinojosa (2020). In this proposal, the maximum and minimum aggregated values are computed for each alternative.

For each unit o ($o = 1, \ldots, n$), the optimistic CI_o^+ and pessimistic CI_o^- evaluations are computed as follows

$$
\begin{aligned}
CI_o^+/CI_o^- = \ max/min \quad & \sum_{r=1}^{m} w_{ro}^+ \cdot I_{ro}^N \\
s.t. \quad & \sum_{i=1}^{n} \sum_{r=1}^{m} w_{ro}^+ \cdot I_{ri}^N = 1 \\
& L_r \leq \frac{w_{or}^+ \cdot I_{ro}^N}{\sum_{r=1}^{m} w_{ro}^+ \cdot I_{ro}^N} \leq U_r, \quad \forall r \\
& w_{ro}^+ \geq 0, \qquad\qquad\qquad \forall r, i;
\end{aligned}
\tag{2}
$$

where L_r and U_r are the lower and upper bounds respectively imposed for the determination of the optimal values of w_{ro}^+ and w_{ro}^-. The outcome from the models (2) is a pair of optimal weighting vectors for each unit (one in the optimistic perspective and a vector for the pessimistic perspective).

To mitigate the effect of outliers and/or the existence of errors, both models are robustified using the concepts proposed in Cazals et al. (2002). DEA models (like other non-parametric models) are highly influenced by the existence of observations with extreme values or by outliers. It is important to bear in mind that the optimal evaluation of each unit is affected not only by its own observation but also by the observations of the remaining units. The procedure proposed in Cazals et al. (2002), suggests computing a number or rounds (2000 in our case) of each model with a sub-sample of randomly selected units and comparing the results to detect the existence of extreme values.

It is important to remark that an alternative DEA-inspired model is proposed in order to determine the optimal values for w_{ri}^+ and w_{ri}^-. In contrast with classic DEA models, not only is the unit under evaluation considered for the construction of the normalisation constraint. In this proposal, the complete set of alternatives participates in the construction of the normalisation condition. The main benefit of this new proposal is derived from the uniqueness of the solution (see Khodabakhshi and Aryavash ([2012](#)) for a detailed explanation).

Following the ideas proposed by Van Puyenbroeck and Rogge ([2017](#)) and Verbunt and Rogge ([2018](#)), we proceed to determine the normalised optimistic and pessimistic weights, denoted respectively by ω_{ri}^+ and ω_{ri}^-. These values are obtained as:

$$\omega_{ri}^+ = \frac{w_{ri}^+ I_{ri}^N}{\sum_{r=1}^m w_{ri}^+ I_{ri}^N}; \omega_{ri}^- = \frac{w_{ri}^- I_{ri}^N}{\sum_{r=1}^m w_{ri}^- I_{ri}^N}. \tag{3}$$

This last step involves the construction of the optimistic and pessimistic geometric indicators. To this end, the original values of the sub-indicators are retrieved. In this phase, a benchmark value or baseline sub-indicator value I_{rB} is considered for each sub-indicator r ($r = 1, \ldots, m$). In this study, the averages of the observed values have been considered as the baseline.

For each alternative, CI_i^+ and CI_i^- are computed as

$$CI_i^+ = \prod_{r=1}^m \left(\frac{I_{ri}}{I_{rB}}\right)^{\omega_{ri}^+}, CI_i^- = \prod_{r=1}^m \left(\frac{I_{ri}}{I_{rB}}\right)^{\omega_{ri}^-}. \tag{4}$$

It is interesting to note that, in ([4](#)), values ω_{ri}^+ and ω_{ri}^- correspond to the relative contribution of sub-indicator r to the aggregated values CI_i^+ and CI_i^-. That is, these values represent the percentage variation in the CI-value as a result of a 1% increase in $\frac{I_{ri}}{I_{rB}}$.

Once the optimistic and pessimistic geometric measures (CI_i^+ and CI_i^-) are computed, both measures can be added to determine a single indicator such that

$$CI_i = \sqrt{CI_i^+ \times CI_i^-} = \prod_{r=1}^m \left(\frac{I_{ri}}{I_{rB}}\right)^{w_{ri}^*} \tag{5}$$

where $\omega_{ri}^* = \frac{\omega_{ri}^+ + \omega_{ri}^-}{2}$.

2.1 Comparing two Subsets of Alternatives

The form of geometric composite indicator proposed above enables the analysis of subsets of indicators, alternatives, or even a dynamic analysis to be carried out. In Verbunt and Rogge ([2018](#)), an analysis of the temporal decomposition of the geometric indicator is proposed. In this work, a comparison between two separate subgroups is realised which, enables a comparative analysis of male (M) and female (F) results. The set of all the alternatives is denoted as $C = \{1, 2, \ldots, n\}$, and consider that the n alternatives are distributed into two separate groups M and F, that is, sets M and F are such that $M \cup F = C$ and $M \cap F = \varnothing$.

In order to perform the comparison, the notation of all the relevant variables should to be extended accordingly in order to include the sub-group reference. Hence the values of the sub-indicators are denoted by I_{ri}^M and I_{ri}^F, by I_{rB}^M and I_{rB}^F the baseline values are denoted,

and the relative importance of the sub-indicators are denoted by by $\omega_{ri,M}^*$ and $\omega_{ri,F}^*$, in all the cases, for subgroups M and F respectively.

If the results of both sub-groups are evaluated separately, then a measure of the performance change for unit i denoted by PC_i, can be measured as follows:

$$PC_i = \frac{\prod_{r=1}^{m} \left(\frac{I_{ri}^F}{I_{rB}^F} \right)^{\omega_{ri,F}^*}}{\prod_{r=1}^{m} \left(\frac{I_{ri}^M}{I_{rB}^M} \right)^{\omega_{ri,M}^*}}. \tag{6}$$

The interpretation of PC_i is straightforward. A value of PC_i greater (less) than the unity indicates that unit i has a better (worse) evaluation for subgroup F than for sub-group M. Note that this interpretation could be carried out separately, by considering optimal weights from optimistic and pessimistic evaluations and the conjoin evaluation.

In Verbunt and Rogge (2018), a tripartite decomposition of PC_i is proposed for the comparison of two successive periods. This decomposition can be extended for the comparison of subgroups such that

$$PC_i = \Delta OWN_i \times \Delta BP_i \times \Delta W_i^*. \tag{7}$$

The component ΔOWN_i measures the changes derived from the variations in the sub-indicators of unit i.

$$\Delta OWN_i = \prod_{r=1}^{m} \left(\frac{I_{ri}^F}{I_{ri}^M} \right)^{\frac{\omega_{ri,F}^* + \omega_{ri,M}^*}{2}}. \tag{8}$$

A value greater (less) than the unity represents an improvement (deterioration) in the performance of the individual indicators in subgroup F with respect to M. That is, a value greater (less) than 1 indicates that the valuation of the indicators I_{ri}, with the corresponding weighting vectors, is greater (less) in F than in M.

With ΔBP_i, the changes derived from the variation of the base-line of F over M are measured.

$$\Delta BP_i = \prod_{r=1}^{m} \left(\frac{I_{rB}^M}{I_{rB}^F} \right)^{\frac{\omega_{ri,F}^* + \omega_{ri,M}^*}{2}}. \tag{9}$$

Here, a value greater than the unity indicates that the composite value of the baseline of M is lower than the corresponding value of subgroup F. Note that, since the sub-indicators are computed in relative terms with respect to the baseline ($\frac{I_{ri}}{I_{rB}}$), a lower value of the baseline value implies an indirect gain to the evaluation.

Finally, the value of ΔW_i^* evaluates the impact of the weighting scheme (comparing whether the evaluation is carried out in the first or the second subgroup).

$$\Delta W_i^* = \frac{\prod_{r=1}^{m} \left(\sqrt{\frac{I_{ri}^M}{I_{rB}^M} \cdot \frac{I_{ri}^F}{I_{rB}^F}} \right)^{\omega_{ri,F}^*}}{\prod_{r=1}^{m} \left(\sqrt{\frac{I_{ri}^M}{I_{rB}^M} \cdot \frac{I_{ri}^F}{I_{rB}^F}} \right)^{\omega_{ri,M}^*}} \tag{10}$$

A value ΔW_i^* greater (less) than the unity indicates that the weighting scheme has been selected in such a way that it represents an advantage (disadvantage) for sub-group F. That is to say, a value greater (less) than the unity indicates a positive (negative) impact derived from the selection of weights in the construction of the composite indicator in subgroup F with respect to M.

3 Case study: Evaluation of the Educational Systems of OECD Countries

The aim of this section is to construct a composite indicator to evaluate the educational system in OECD countries. The proposal involves the application of the methodology described in the previous section to the values included in the Programme for International Student Assessment (PISA) report for 2018 (OECD, 2021a). The PISA report, initially launched in 2000, offers triennial statistics on the structure and performance of educational systems. The PISA database constitutes a major source of information for the development of comparative analysis across economies. The number of participating schools and countries has risen in every edition of the report, currently standing at about 500,000 students from 80 countries (in the edition of 2018).

Even though the main target of the study is to evaluate the skills and knowledge of 15-year-old students in Mathematics, Science, and Reading skills and, since 2012, financial literacy has also been included as an option, the database includes a large amount of interesting data related to academic achievements (results from standardised test scores), the students' households, and on the schools they attend. Furthermore, it contains synthetic indices created by OECD experts of some interesting aspects (see OECD (2021a) for a detailed discussion).

The results for the complete set of students are obtained for every country (the complete set denoted by C in the previous section) as are the separate results for female and male students (denoted by F and M, respectively), in order to carry out a comparison of these two subgroups. The multiplicative nature of the proposed methodology permits the results of these comparative results to be separated into three effects, and the contribution of each considered dimension can therefore be determined.

3.1 Panel of Indicators

In this paper, we include a proposal of a panel of individual indicators for the comprehensive evaluation of educational systems that extends beyond the simple computation of academic outcomes. An expansion of the concept of the objective of a educational system is proposed. We consider that, although the main objective to achieve is the optimal transfer of skills to the students, a transforming role in a social context must also constitute one of the desirable goals of the systems. In other words, the educational function should be broadly interpreted by including social features and the well-being of the students during their scholar stage.

Therefore, in the evaluation of the performance and outcomes from the educational systems, not only are the academic results taken into account.

We classify the sub-indicators into three main areas or dimensions in order to cover the entire set of valuable elements: academic outcomes, social equity, and the students' perception of the system. The following scheme summarises the main concepts studied in each dimension and the selected subindicators (in italics).

(a) *Academic outcomes*. In the first dimension, we propose the inclusion of the academic results obtained by the students. It is clear that the better the academic results, measured by results from normalised test scores, the higher the standard of the educational system. On the basis of the quantification of the education received by an individual this is not an easy task, due to its inherent intangibility and the necessity of considering a long period. However, there is certain consensus in the consideration of standardised test results as one of the main educational outcomes. In Morrison (2011), a complete review of the main aspect related with standardised tests in the OECD context can be seen, including the implication for the students in their future academic outcomes.

This dimension includes a total of five indicators.

 (a.1) *Academic average performance*. With the academic average performance, the overall achievement of the students is measured. The average student learning outcome is considered through the average of the standardised test scores in the main disciplines included in PISA: Mathematics (*MATH*), Science (*SCI*), and Reading (*READ*) in order to compute the average outcomes of education as a production process.

 (a.2) *Excellence of the educational system.* A high-quality education is desirable in any modern society. In order to take this aspect into account, the students' proficiency levels published in PISA reports have been examined. The results from PISA tests are represented by means of six levels of educational proficiency, built from the test scores (a detailed explanation is given in OECD (2017)). PISA reports establish that an educational system achieves the minimum objectives if the students achieve at least the second level of proficiency. Students who reach levels equal to or greater than 5, are considered to present high performance. In order to take into account the excellence of the educational system, an indicator (*EXCE*) is considered that computes the percentage of students who obtain a level equal to or greater than 5 in any of the three main subjects measured in PISA.

 (a.3) *Academic inclusion.* One of the desirable objectives of educational systems is that all students reach at least a minimum level of knowledge (defined as a baseline level of skills). We consider an indicator (*INCL*) that measures the percentage of students who reach at least second proficiency level in the three referred subjects. It is reasonable to assume that one of the goals of the system is to guarantee that all the students reach at least a baseline level of skills in all subjects.

(b) *Equity of the educational system.* The second aspect to be studied in this multidimensional indicator involves the equity derived from the performance of the educational system. The transforming role of the educational system is evaluated when access to learning goes beyond the socio-economic background of the student so that the required mechanisms can guarantee that every interested student has the opportunity to attain his/her academic achievements. Equity in education constitutes a specific target of the Sustainable Development Goals set by the United Nations in 2015. Equity does not mean that all students will have the same results in education, but that their results will not be conditioned by their circumstances (Downey & Condron, 2016; Roemer & Trannoy, 2015).

Several meta-analyses related to socioeconomic status and students' academic performance can be found in the literature (e.g., Liu et al., 2019; Selvitopu & Kaya, 2021; Sirin, 2005; White, 1982). The main results pointed out that the relation between socioeconomic status and academic performance is, in general, positive but depends on the

particular measurement considered to represent socio-economic achievement. In this paper, two approaches to this concept are considered: social equity and socioeconomic incidence.

(b.1) *Resilient students.* Ideally individuals should be able to obtain excellent academic results depending on only their individual abilities, no matter how adverse the conditions in their environment may be. The percentage of *resilient students* is included as an indicator (*RESI*). In PISA, a student is classified as resilient if he or she is placed in the bottom quarter of the PISA index of Economic, Social and Cultural Status (ESCS) and placed in the top quarter of students regarding academic results. At this point, the definition given by Agasisti et al. (2018) has been borne in mind.

(b.2) *Socio-economic incidence.* The equity in an educational system can also be measured by the existence of a favourable context that permits all the students to develop their talent, and to overcome the limitations derived from their economic and social circumstances. It is clear that the quantification of this idea is not straightforward. We try to approximate this point through the percentage of variation in the scores of the main subjects that cannot be explained by a student's socio-economic status. Three different indicators have been considered to measure the incidence in Science (*INCS*), Mathematics (*INCM*) and Reading (*INCR*).

The idea here is to measure the capacity of a system in minimising the influence of socio-economic background on the academic achievements of the students. It is interesting to note the main differences between the two groups of indicators included in this dimension. Note that the concept of resilient students only considers those students that are placed in the bottom quarter of the ESCS. In contrast, the former indicator takes into account the complete set of students.

(c) *Students' self-perceptions.* One of the main characteristics of the education function Mancebon and Bandres (1999) is, among others, that the educational process is carried out by the customers themselves. This justifies the incorporation of the self-evaluation of the student into the complete evaluation of the educational system. Hence, a third dimension is included in an effort to measure the students' well-being during their scholar stage. Although the desirability of excellent academic results is paramount, this target should not be achieved at a cost to the students' well-being. Hence, we propose the inclusion of a group of indicators that reflect the self-perceptions of the students during their scholar stage.

Previous studies suggested that the school environment is a key aspect to understand the degree of satisfaction of the students with teachers, classmates... (e.g., Casas et al., 2013; Rees & Main, 2015). In Berkowitz et al. (2017) there is a comprehensive literature review of those studies that examined if a positive climate at the school has a positive influence on academic outcomes and can mitigate the association between socio-economic status and academic achievements. From the different results, the direction and relations between socio-economic status, school climate, and academic performance is not unique and conclusive.

Some previous proposals (see, for instance, Rothstein, 2000) suggested a composite index of school performance that included aspects such as students' security at school, the adult attention the students receive and the role of teachers in addition to academic results.

We strive measure these self-perceptions with respect to the environment of the educational institutions, their wellness, the teachers' performance, and the added value of the scholar ages for their future working lives.

(c.1) *Sense-of belonging.* Firstly, we include one of the composite indices developed by PISA as a proxy of the overall well-being of the students at school: the *Sense of Belonging* index. This indicator (*BELO*) attempts to measure how accepted, respected, and supported students feel in their social context at school (Goodenow & Grady, 1993). Previous studies have shown a positive association of this variable with other, such as positive disciplinary climate, participating in extracurricular activities, family support, and positive teacher-student relations.

(c.2) *Bullying exposure.* A second aspect of interest is that of the prevention of bullying behaviour. This aspect is currently one a growing scourge in almost all countries. The composite indicator developed by PISA, *Index of Exposure to Bullying*, is included, which quantifies the exposure to bullying with respect to average student across countries (*BULL*).

(c.3) *Motivation in teaching goals.* A third relevant aspect for the students involves the motivation in their day-to-day activities. This is considered to be a positive aspect of an educational system if the environment of the schools and the motivation of the students are such that the students' goal of their daily activity is to learn as much as possible or to understand the content of the classes. We have included the composite index proposed by PISA to measure the ambition towards the *L*earning Goals with respect to the average (*GOAL*).

(c.4) *Valuation of the teachers' activity.* It is clear that one of the most determining elements in the education process is that of the role of teachers. We propose the inclusion of the students' evaluations regarding the teachers' attitude (*TEACH*). We have included the index of *Teacher Support* proposed in PISA, which strives to measure the interest level of the teachers in the learning process or the extent to which the teachers help.

(c.5) *Valuation of the scholar stage.* Finally, the inclusion of the index *Value of School* proposed in PISA strives to measure how the students quantify the added value that schools incorporate into their subsequent scholar stages and their future working lives. This index (*V ALU*) is developed by including the answers to a set of questions such as whether trying hard at school will help towards entering a good university and/or to get a good job in the future.

To summarise, a list of fourteen indicators, separated into the aforementioned three dimensions, is considered. Table 1 outlines the main descriptive statistics for the set of indicators.

Table 2 summarises the results of independent samples *t*-test and proportion test for comparison of the proposed indicators with respect to gender both globally and segmented by each of the 37 OECD countries included in this work. Note that only eleven of the fourteen variables have been tested, since the three variables extracted from the incidence of ESCS index could not be compared with the respect to the average values. In those cases in which the variables are constructed from plausible values, a separate analysis of the variables has been carried out, taking into account each plausible value. In all the cases, the results have been weighted with the size of the sample. For each country, the normalised weights are used. In this way, comparisons between countries can be performed, thereby obtaining robust estimations.

Significant differences between men and women are found in all the studied variables, except *TEACH* and *SCI*. Regarding average scores in Mathematics, the performance by men is better than that of women, the opposite being the case in Reading. It is interesting to point out that the differences in the latter case (*READ*) are greater. Similarly, men have a higher average value in the index of *BELO* and women in *BULL* (note that this indicator must be modified so that a higher value indicates a better situation).

The highest number of differences are found in Finland, Norway (with differences in all the variables but one, index *TEACH*), Canada (in which all the differences appear in the same direction of global evaluation), and Israel (in which no differences exist with respect to *EXCE*). In contrast, USA is the country in which a minimum number of differences are found between men and women. The only differences arise in *MATH* and *BELO*, in favour to men, and in *READ*, *GOAL*, and *VALU*; in which women perform better than men.

With respect to each individual indicator, women perform better than men with respect to variables *READ*, *VALU*, and *GOAL*. The opposite situations appear in *BELO*, in which the higher performance corresponds to the group of men. It is interesting to highlight the case of *RESI*. Significant differences appear in favour of men in the case of Chile, Colombia, Spain, and Mexico. The case of Spain may be justified since the information for Reading is not available, in which women have generally better results than men.

It is interesting to highlight how the third dimension presents a negative or statistically non-significant correlation with the other indicators. Concerning this point authors such as Mikk et al. (2016) and Ma et al. (2021) have shown how the effect of student-perceived teacher support on reading literacy is not significant at the school level, but it is at the student level. In addition, the correlation between teacher-students relations and the academic outcome shows a weak positive relationship at the student level, a positive correlation at the school level but a negative correlation at the country level.

3.2 Discussion of the Results

The results obtained for the composite indicator are summarised in Table 3. The table includes the results obtained for each country in three cases: global (considering all the students), male and female (considering separately the results of each subgroup of students). In addition, the ranking induced by these values has been also included. Note that in this order, the first position is assigned to the top-performing country and 37th position to the worst-rated country, and that each ranking is constructed by considering solely the values of the corresponding subgroup, since this indicator, as seen in the section of methodology, must be analysed with respect to the set of alternatives under consideration in each evaluation.

For illustrative purposes, the results of a weighted additive aggregation have been also included in Table 3. It is important to note that this aggregation scheme permits a complete compensation between the indicators. Although the majority of the results are similar, some units take advantage of this feature. That is the case of Slovenia, which is ranked in the 30th position in the geometric indicator and in the 25th in the additive aggregation. This unit presents really good values in three indicators (*READ*, *INCL* and *RESI*) but poor values in the remaining ones. This circumstance, an unbalanced vector of observations, is reflected in a poor evaluation when a geometric aggregation is considered. But the full compensation between sub-indicators is allowed (additive aggregation) having as a consequence a better evaluation for the unit.

It is important to bear in mind that each indicator has been constructed with an individual vector of weights and that an upper bound of 0.15 and a lower bound of 0.03 have been

included in the relative importance of the weights. These vectors are computed taking into account the observed values of the remaining units of the subset (and not the global evaluation of both subgroups). That is, for the determination of the values of male indicators, only the values of the sub-indicators of this subgroup were considered for the unit under evaluation and for the remaining units. By way of example, Tables 4 and 5 (see Annex) include the normalised weights for the global data in the optimistic and pessimistic evaluation, respectively. In is interesting to bear in mind that the model which determines optimal weights proposes a relative evaluation of the units. That is to say, in the optimistic evaluations, the optimal weights will be selected to identify which variables the unit presents a comparative advantage with regarding the remaining ones (by giving higher values). In contrast, in the pessimistic evaluations, those in which the composite evaluation of each unit is minimised, comparative disadvantages are identified by assigning higher values.

The results of the rankings are represented for comparison purposes in Fig. 1. For each country, the positions occupied in the orders induced in the global evaluation and the evaluations of each subgroup (male and female students) are represented. When the rankings coincide, the markers are overlapped. Otherwise, the vertical distance between points measures the difference between the position in each order. In general terms, there are no large differences between the three orders, with two exceptions in which differences between rankings attain seven units: ESP, with a better performance of the male subgroup, and ISL, where the worst evaluation corresponds to male students. The rank position induced in global and male evaluations coincide in 15 times, and the order position in the global ranking and in the female ranking is the same 28 times.

The multiplicative nature of the CI permits enables the identification of the individual contribution of each of the three dimensions considered. In Table 6 (see the Appendix), the values of the indicator separated by subgroups and each dimension are included. Figure 2 presents the contribution of each dimension to the global case (similar figures could be constructed for male and female evaluations). The solid bars represent the value of the weighted indicator for each dimension (academic, equity, and students' self-perceptions). The height of
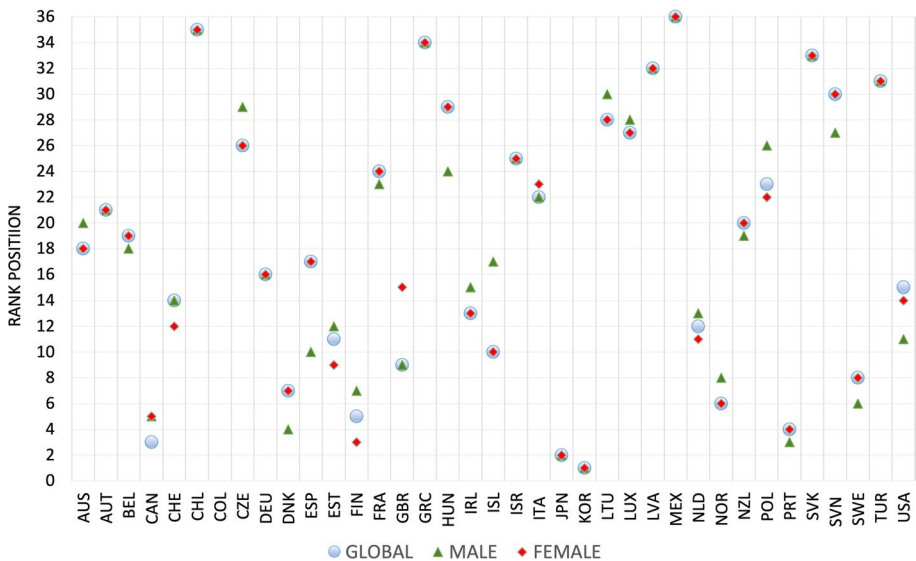
Fig. 1 Induced rankings. Comparative results
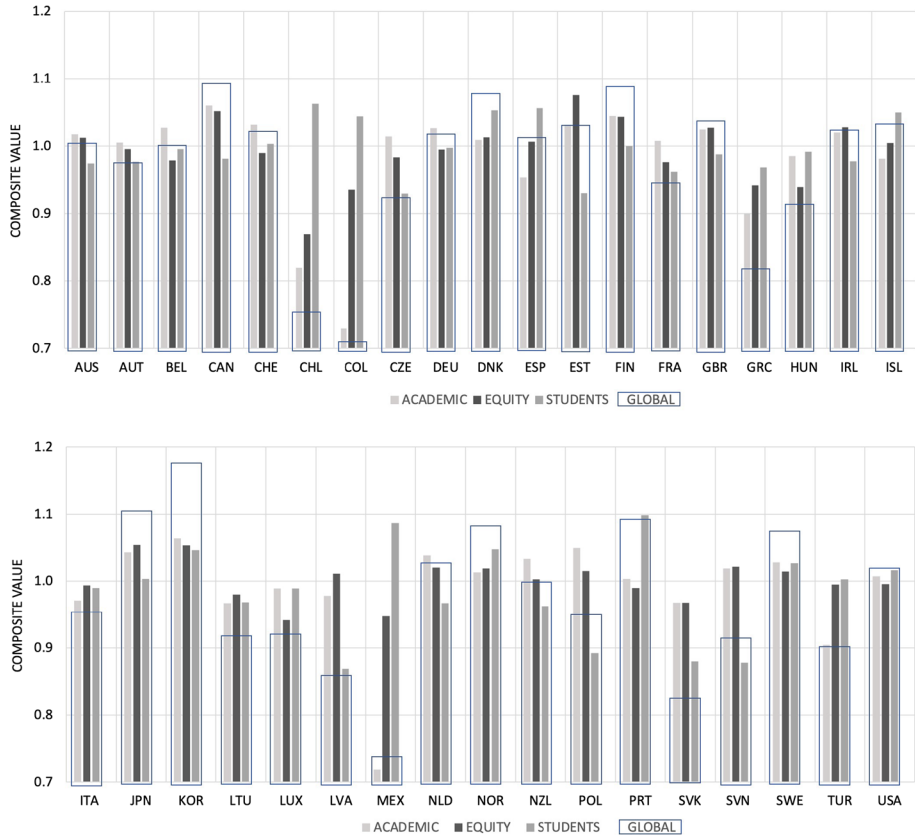
Fig. 2 Contribution of each dimension to the composite index

the non-filled box represents the value of the CI. This graph permits to detect, the strengths and weakness of each country to be detected (aggregated by the three referred dimensions).

It is interesting to see how, due to the multiplicative nature of the composite indicator, those units with a "well-balanced" situation are rewarded with a greater value of the composite indicator, especially if the value of the components is greater than the unit (the individual indicator is over the average). These are the cases of the best-valued countries: Korea, Japan, Sweden, and Finland.

In the opposite situation, we found countries with a low value in one (or more than one) dimension. The geometric aggregation does not permit compensation between parts and penalises the existence of low values in one of its parts. This is the situation of Mexico and Colombia, where the poor performance in the academic dimension drops the global result even lower. This situation would force the units to focus their efforts on the improvements of those aspects in which the observed values are lowest. A marginal progress in the academic dimension would have a major effect for these countries, much more than if the same improvement occurred in any of the other dimensions. That is to say, the geometric aggregation implies a penalisation factor that captures the unbalance from the observed values. This aggregation scheme in some way permits evaluating more appropriately a complex phenomenon, since all the valuable attributes or aspects are jointly considered.
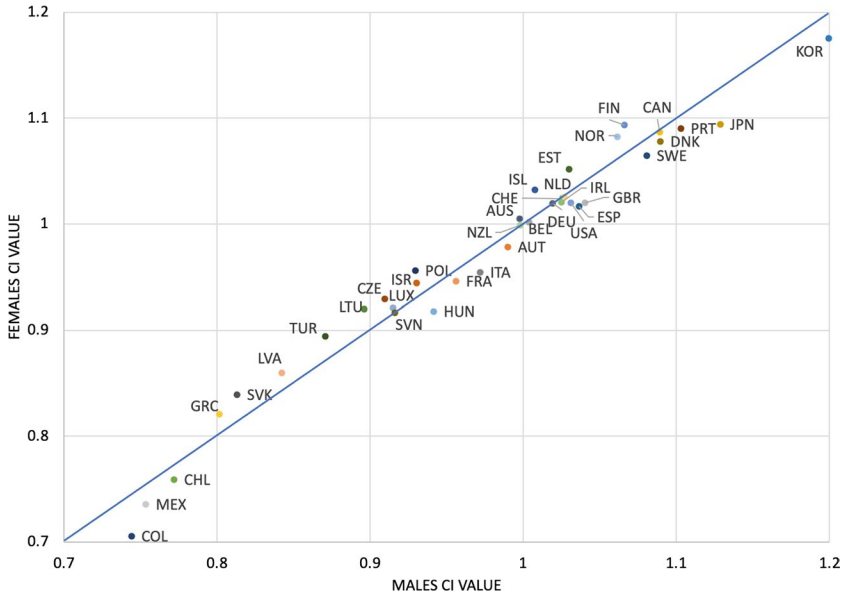
Fig. 3 Global Indicator. Comparative male vs. female

A correlation analysis (Spearman's Rho) enables us to see that the global evaluation, together with the academic and equity dimensions, present similar values, with positive and statistically significant correlation coefficients. This situation appears when two of these sets of values are compared for an individual evaluation (total, male, and female) as well as for an individual dimension (global, academic, and equity).Conversely, the third dimension (students' self-perceptions) presents negative or statistically non-significant correlation coefficients when they are compared with the other three values in both cases: they are compared with different evaluations as well as with respect to the other three values of the same group of students.

Figure 3 enables the comparison between countries considering the subgroups' evaluations. Each point represents the values of the composite indicator for males (abscissa axis) and for females (ordinate axis). Those units located at a distance from the origin represent the countries with a better evaluation in both subgroups (Korea and Japan). Conversely, those countries closer to the origin represent the worst performers (Colombia, Mexico, and Chile).

The distance from the origin to the right (upwards) indicates a better valued unit in the male (female) indicator. A continuous line that represents the bisector has also been incorporated. Those units located above this line are those in which the index in the female subgroup is greater than the value in the male index. The units located below the line present the inverse situation. If the indicator had coincided for the two subgroups, then the unit would be located on the bisector. Note that the larger the difference of angle with respect to 45°, the greater the differences of the evaluation of both subsets.

It is interesting to see that, 17 times, the numeric value of CI is greater for the female evaluation (and, conversely, in 20 countries the value of CI is greater for the male subgroup). The greatest differences in favour of the male subgroup are found in Japan and Colombia. The greatest differences where the CI of the female subgroup is higher correspond to Poland and Finland.
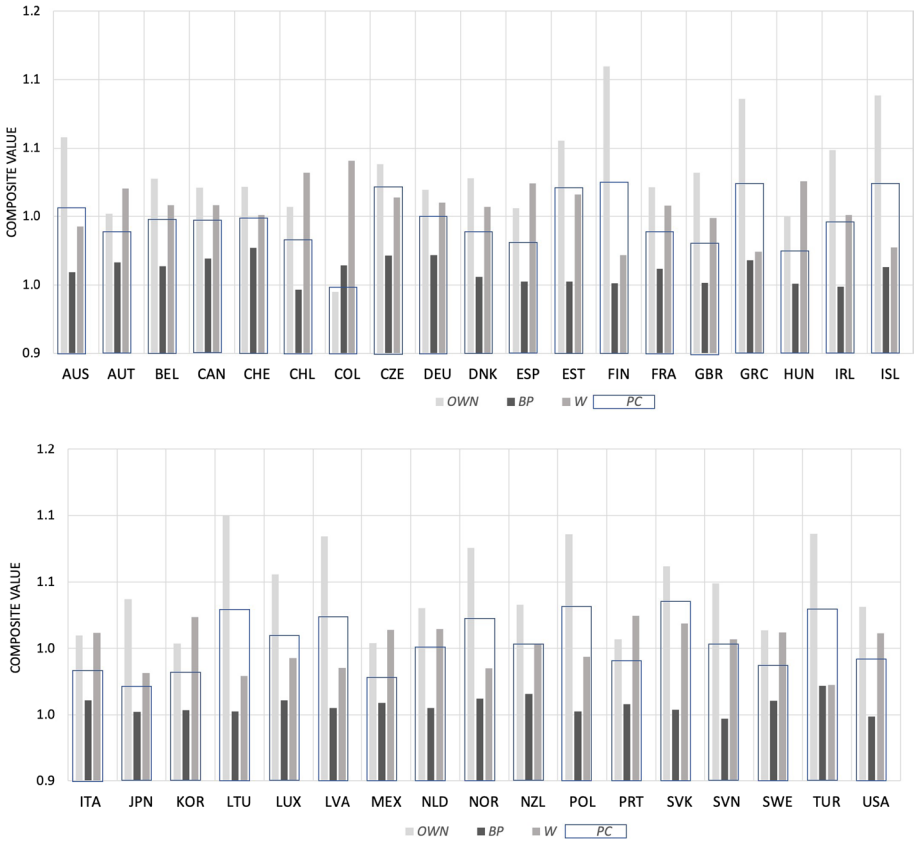
Fig. 4 Decomposition of the comparison between female and male results

Finally, Table 7 (see Appendix) summarises the decomposition of the values PC_i, which enables the identification of the main causes that explain a greater value of CI_i for one of the subgroups.

In a similar way to that in the Fig. 4, the filled bars represent the value of the individual effects (own performance, base, and weighting vectors) while the height of the non-filled box represents the value of PC_i which compares the composite indicator for females over the corresponding value for males. In this way, a value of PC_i greater than the unity implies that, for this country, the value of the CI is greater for the female students than for the male students.

The values of *OWN* represent the contribution (positive or negative) of the observed sub-indicator. Values greater than the unity represent a better result in the country derived exclusively from a better performance in the individual indicators. All the countries except two (Colombia and Hungary) present values greater than the unity in this component, with special mention to Finland and Lithuania, which present the largest differences in favour of the female subgroup.

The *BP* values quantify the influence of the variation of the baseline in the composite indicator. Here, the comparison of the baseline of male over female is measured. Note that, in this case, greater values of the bases would have a negative influence on the composite indicator since the value of the indicator is included through $\left(\frac{I_{ri}}{I_{rB}} \right)$. That is, a lower value of

the baseline exerts an indirect positive effect on the composite indicator. Values lower than the unity (common in all countries) reflect that the baseline for the female subgroup is greater than the corresponding values for the male subgroup.

Finally, W enables the incidence of the weighting vectors over PC_i to be measured. A value greater (less) than the unity indicates that the selection of the weighting factor in female induces an improvement (deterioration) in values of this subgroup with respect to those of the male subgroup. In this case, the results vary although they remain favourable for female students, since the effect of the selection of optimal weights is positive for the evaluation of this subgroup (values greater than the unity) in 24 countries.

4 Concluding Remarks

Composite indicators have proved in the past to be suitable for the evaluation of complex phenomena. In this work, we propose a methodology for the construction of composite indicators based on the multiplicative aggregation and free selection of weights. In brief, a two-step procedure based on Data Envelopment Analysis in the determination of the weighting profiles has been described in order to construct a composite indicator. The main objective is to develop a valuable tool for the comparison of the performance of educational systems. We are interested in providing a global measure that yields a global evaluation for each country, which enables the strengths and weakness of the evaluations to be identified, and comparisons to be performed between the results obtained by male and female students separately. The main contribution is the development of an aggregation scheme that permits identifying the sources of relative advantages or disadvantages of one unit or subgroup with respect to the others.

A geometric aggregation of the sub-indicators in order to avoid the compensation between individual aspects has been considered. This family of indicators reward those alternatives with a well-balanced situation (above average values of all the indicators) over those alternatives with a higher value in one or two of the aspects, but very low results in one of the remaining aspects. Furthermore, the multiplicative nature of the indicator enables the easy identification of the contribution of each dimension to the global value. It was observed that the worst-ranked countries correspond to those with a very low values in any of the referred dimensions (academic, such as Colombia and Mexico, or students' self-perceptions, such as Slovenia and Latvia) that cannot be compensated by good results in the other aspects. This feature of the indicators based on geometric aggregation provides information the countries as to which aspects must be improved, since improvements in worst-valued aspects will have a greater impact on the composite indicator. To certain extent, this provides a first approach to the general outlines for political measures. On the other hand, if the sources of the low values are the weighting schemes or the basis considered as a benchmark, the aim of the political measures should be a global improvement of the observations.

Herein, the results of male and female students have been compared separately. Considering them as two independent populations, we computed the composite indicators for both subgroups and compared the results. It is important to point out that this analysis implies computing all the results for the two subgroups separately. That is to say, the set of individual sub-indicators, the baseline that provides a basis for the normalisation phase, and the optimal weighting profiles are different for the first evaluation (which includes all the observations), to those for the other two computations.

It is interesting to bear in mind that the aim of this paper is not to perform a gender-perspective study like those found in the literature on this topic. Our intention is not to

measure the gender-gap, to identify the sources of the differences between the values, or to propose political measures to mitigate these differences. Instead, our approach strives to measure the performance of national educational systems by considering a global view, and to evaluate separately the two groups of students in an effort to point out those cases (countries) where the results are significantly different. These results can be seen as a starting point for a posterior gender perspective analysis, in order to identify the causes of these differences and to propose measures for the mitigation of these gaps.

A first approach to the individual sub-indicators gave a first comparison between female and male results. Clearly, the male results in Mathematics and Sense of Belonging are better than the female results; the opposite occurs in Reading and certain indicators included in the third dimension (Motivation in teaching goals). When all these individual values are gathered together as the components of the composite indicators, the results indicate that there are no significant differences in the global indicator or in any of the dimensions regarding gender (results that are confirmed through Mann–Whitney tests).

When the aggregated evaluations are analysed, a similar performance for male and female results can be observed. In fact, a high correlation exists between the results of the two subgroups and between the results of two of the three dimensions. Only the third dimension, that which measures the self-perception of the students, presents a different performance with respect to the global or individual evaluation.

The proposed methodology permits to analyse main sources of the differences between male and female results. The proposed model enables the identification of whether a better result of one of the subgroups (male or female) is derived from its own observations, the impact of the baseline, or the optimal weighting profiles. On this point, it is interesting to see how the higher value of the average in the female subgroup exerts, in general terms, a negative impact on the evaluation since a higher value of the observed sub-indicators is required to lead the normalised value to be greater than the unity.

It would also be of interest to analyse the results with respect to the main economic variables traditionally associated with education. In particular, we consider GNP and accumulative educational expenditure per student. The correlation between these two variables is high (0.881) and the values of the composite indicator are statistically significant. Likewise, when the ranges in these variables and in the global indicators are considered, it can be observed that the correlations between the global indicators, those of dimension 1 (academic) and 2 (equity) (both globally and segmented by gender), are statistically significant. Nevertheless, the results for third dimension (students' self-perceptions) do not depend on either the academic or the social part. In this case, the correlation is negative or non-significant, that is, countries with a low value for GNP and accumulative educational expenditure can achieve good results in this third dimension, perhaps due to the work of heads teachers and teachers, which overcome the difficulties derived from a lack of resources.

Appendix

See Tables 1, 2, 3, 4, 5, 6 and 7.

Table 1 Descriptive statistics

Indicator	Global				Male				Female			
	Min	Max	Average	Std. Dev	Min	Max	Average	Std. Dev	Min	Max	Average	Std. Dev
SCI	413.32	530.11	488.66	26.25	419.61	530.56	487.52	25.68	407.29	534.04	489.81	27.28
MATH	412.30	523.02	487.12	24.87	407.03	507.70	472.38	24.03	417.36	546.34	502.08	26.62
READ	390.93	526.97	489.29	30.46	400.89	532.16	491.87	29.80	381.37	523.81	486.67	31.37
EXCE	1.11	26.56	15.50	6.08	1.18	27.02	15.64	6.02	1.04	26.06	15.36	6.31
INCL	28.84	83.20	66.79	11.36	32.55	81.10	64.56	10.50	25.28	85.31	69.06	12.46
RESI	2.37	43.28	22.69	9.13	3.34	42.55	22.06	8.73	1.62	43.98	23.31	9.72
INCS	78.83	93.58	87.18	3.96	81.55	94.68	88.51	3.54	75.44	92.87	85.59	4.60
INCM	80.94	93.80	87.99	3.50	83.87	95.29	89.20	3.02	76.42	92.40	85.91	4.36
INCR	76.18	92.18	86.16	3.93	79.19	93.84	87.47	3.50	72.98	91.36	84.57	4.61
BELO	0.72	1.46	1.00	0.20	0.67	1.44	1.00	0.21	0.73	1.49	1.00	0.20
BULL	0.59	1.29	0.99	0.16	0.48	1.23	0.91	0.18	0.71	1.38	1.08	0.16
GOAL	0.65	1.54	1.00	0.23	0.58	1.46	0.92	0.21	0.71	1.64	1.11	0.26
TEACH	0.54	1.40	1.00	0.19	0.46	1.27	0.91	0.19	0.62	1.52	1.09	0.19
VALU	0.38	1.46	1.00	0.25	0.44	1.42	1.00	0.24	0.31	1.51	1.00	0.27

Table 2 Results of independent samples t-test

	MATH	READ	SCI	EXCE	INCL	RESI	BELO	BULL	GOAL	VALU	TEACH
Australia	AUS	M**	F**	–	F*	F**	F*	M**	F**	F**	F**	–
Austria	AUT	M**	F**	–	M**	F**	F*	–	F**	F**	F**	M**
Belgium	BEL	M**	F**	–	M**	F**	F**	M*	F**	F**	F**	–
Canada	CAN	M**	F**	–	M*	F**	F**	M**	F**	F**	F**	n.a
Switzerland	CHE	M**	F**	–	–	F**	F**	M*	F**	F**	F**	M**
Chile	CHL	M**	F**	–	–	–	M**	M**	F**	F**	F**	–
Colombia	COL	M**	F**	M**	–	M**	M**	M**	F**	F**	–	–
Czech Republic	CZE	–	F**	–	–	F**	F**	M**	F**	F**	F**	–
Germany	DEU	M**	F**	–	–	F**	F*	–	M*	F**	F**	M**
Denmark	DNK	M*	F**	–	–	F**	F*	M**	F**	F**	F**	M**
Spain	ESP	M**	n.a	–	M**	–	M**	M*	F**	F**	F**	F**
Estonia	EST	M**	F**	–	–	F**	F**	M**	F**	F**	F**	–
Finland	FIN	F**	F**	F**	F**	F**	F**	M**	F**	F**	F**	–
France	FRA	M**	F**	–	M**	F**	F*	M*	–	F**	F**	–
United Kingdom	GBR	M**	F**	–	–	–	–	M**	F**	F**	F**	–
Greece	GRC	–	F**	F**	–	F**	F**	–	F**	F**	F*	–
Hungary	HUN	M**	F**	–	–	M*	–	M**	F**	–	F**	M**
Ireland	IRL	M*	F**	–	–	F**	–	M**	F**	F**	F**	M**
Iceland	ISL	F**	F**	F*	–	F**	F**	M**	M*	F**	F**	–
Israel	ISR	F**	F**	F**	–	F**	F**	n.a	n.a	F**	F**	M*
Italy	ITA	M**	F**	–	M**	–	–	M**	F**	F**	F**	–
Japan	JPN	M**	F**	–	M**	F**	F**	–	F**	–	F**	–
Korea	KOR	–	F**	–	–	F**	F**	–	n.a	–	–	M**
Lithuania	LTU	–	F**	F*	–	F**	F**	M**	F**	F**	–	–
Luxembourg	LUX	M**	F**	–	–	F**	F**	–	F**	F**	F**	–
Latvia	LVA	M**	F**	F**	–	F**	F**	–	F**	F**	F**	M**
Mexico	MEX	M**	F**	M**	–	M**	M**	–	F**	F**	F**	F*

Table 2 (continued)

		MATH	READ	SCI	EXCE	INCL	RESI	BELO	BULL	GOAL	VALU	TEACH
Netherlands	NLD	–	F**	F*	–	F**	F**	–	F**	F**	F**	M**
Norway	NOR	F**	F**	F**	F**	F**	F**	M**	F**	F**	F**	–
New Zealand	NZL	M**	F**	–	–	F**	–	M**	F**	F**	F**	–
Poland	POL	–	F**	–	–	F**	F**	M**	F**	F**	F**	F**
Portugal	PRT	M**	F**	–	M**	F**	–	M**	F**	F**	F**	F**
Slovak Republic	SVK	–	F**	–	–	F**	F**	–	F**	F**	F**	M**
Slovenia	SVN	–	F**	F**	–	F**	F**	–	F**	–	F**	M**
Sweden	SWE	–	F**	F*	F**	F**	F**	M**	F**	–	F*	–
Turkey	TUR	M*	F**	F**	–	F*	–	M**	F**	F**	F**	F**
United States	USA	M**	F**	–	–	–	–	M**	–	F**	F**	–
Global		M**	F**	–	M**	F**	F**	M**	F**	F**	F**	–

F** Average in female group greater than at 99% of significance

F* Average in female group greater than at 95% of significance

–No significant differences are found

M** Average in male group greater than at 99% of significance

M* Average in male group greater than at 95% of significance

n.a. Not available data

Table 3 Composite indicators and induced ranking

	Geometric aggregation						Additive aggregation					
	Global		Male		Female		Global		Male		Female	
	CI_I	Rank	CI_I	Rank	CI_I	Rank	CI_I	Rank	CI_I	Rank	CI_I	Rank
AUS	1.0030	18	0.9978	20	1.0052	18	1.0155	19	1.0116	20	1.0183	18
AUT	0.9773	21	0.9900	21	0.9781	21	0.9968	21	1.0063	21	0.9978	21
BEL	1.0008	19	1.0037	18	1.0015	19	1.0076	20	1.0133	19	1.0083	20
CAN	1.0938	3	1.0893	5	1.0867	5	1.1129	3	1.1052	4	1.1038	4
CHE	1.0244	14	1.0257	14	1.0245	12	1.0321	15	1.0329	15	1.0323	14
CHL	0.7573	35	0.7721	35	0.7592	35	0.8737	34	0.8752	34	0.8741	34
COL	0.7123	37	0.7442	37	0.7055	37	0.8599	35	0.8794	33	0.8551	35
CZE	0.9267	26	0.9096	29	0.9295	26	0.9354	27	0.9203	30	0.9378	27
DEU	1.0188	16	1.0192	16	1.0195	16	1.0286	17	1.0276	16	1.0293	16
DNK	1.0768	7	1.0897	4	1.0781	7	1.0834	7	1.0975	5	1.0846	8
ESP	1.0132	17	1.0364	10	1.0169	17	1.0344	13	1.0541	10	1.0379	12
EST	1.0321	11	1.0301	12	1.0518	9	1.0655	9	1.0642	9	1.0889	7
FIN	1.0891	5	1.0662	7	1.0933	3	1.1011	5	1.0714	8	1.1058	3
FRA	0.9464	24	0.9562	23	0.9462	24	0.9513	26	0.9634	23	0.9511	26
GBR	1.0394	9	1.0404	9	1.0202	15	1.0542	11	1.0536	12	1.0327	13
GRC	0.8216	34	0.8017	34	0.8208	34	0.8557	36	0.8357	36	0.8545	36
HUN	0.9178	29	0.9416	24	0.9177	29	0.9289	29	0.9497	26	0.9286	29
IRL	1.0248	13	1.0249	15	1.0205	13	1.0336	14	1.0373	14	1.0290	17
ISL	1.0346	10	1.0078	17	1.0323	10	1.0407	12	1.0168	18	1.0383	11
ISR	0.9433	25	0.9303	25	0.9441	25	0.9616	23	0.9507	25	0.9625	23
ITA	0.9553	22	0.9720	22	0.9543	23	0.9611	24	0.9751	22	0.9599	24
JPN	1.1029	2	1.1286	2	1.0941	2	1.1266	2	1.1544	2	1.1162	2
KOR	1.1735	1	1.1995	1	1.1752	1	1.1928	1	1.2190	1	1.1939	1
LTU	0.9179	28	0.8962	30	0.9196	28	0.9246	31	0.9041	31	0.9265	30
LUX	0.9212	27	0.9150	28	0.9212	27	0.9304	28	0.9252	28	0.9300	28
LVA	0.8593	32	0.8424	32	0.8600	32	0.8777	33	0.8642	35	0.8782	33
MEX	0.7397	36	0.7537	36	0.7356	36	0.9165	32	0.9231	29	0.9144	32
NLD	1.0255	12	1.0269	13	1.0255	11	1.0543	10	1.0539	11	1.0544	10
NOR	1.0813	6	1.0616	8	1.0822	6	1.0906	6	1.0719	7	1.0916	6
NZL	0.9977	20	0.9981	19	0.9987	20	1.0169	18	1.0221	17	1.0179	19
POL	0.9503	23	0.9296	26	0.9558	22	0.9781	22	0.9577	24	0.9852	22
PRT	1.0906	4	1.1030	3	1.0902	4	1.1034	4	1.1158	3	1.1030	5
SVK	0.8244	33	0.8133	33	0.8393	33	0.8345	37	0.8261	37	0.8494	37
SVN	0.9143	30	0.9163	27	0.9168	30	0.9529	25	0.9429	27	0.9553	25
SWE	1.0713	8	1.0808	6	1.0647	8	1.0742	8	1.0827	6	1.0674	9
TUR	0.9030	31	0.8710	31	0.8943	31	0.9288	30	0.8950	32	0.9206	31
USA	1.0190	15	1.0310	11	1.0203	14	1.0298	16	1.0391	13	1.0312	15

Table 4 Normalised weights ω^+ for the global data

	SCI	MATH	READ	EXCE	INCL	RESI	INCS	INCM	INCR	BELO	BULL	GOAL	TEACH	VALU
AUS	0.030	0.058	0.030	0.084	0.030	0.030	0.150	0.120	0.144	0.030	0.030	0.030	0.084	0.150
AUT	0.064	0.030	0.150	0.068	0.147	0.030	0.030	0.030	0.030	0.150	0.138	0.073	0.030	0.030
BEL	0.063	0.030	0.148	0.150	0.033	0.030	0.030	0.030	0.030	0.098	0.150	0.148	0.030	0.030
CAN	0.030	0.030	0.030	0.150	0.030	0.124	0.150	0.150	0.150	0.030	0.030	0.035	0.031	0.030
CHE	0.145	0.030	0.150	0.150	0.135	0.030	0.030	0.030	0.030	0.150	0.030	0.030	0.030	0.030
CHL	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.131	0.030	0.030	0.149	0.150	0.150	0.150
COL	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.128	0.150	0.030	0.030	0.150	0.032	0.150
CZE	0.150	0.130	0.150	0.150	0.150	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030
DEU	0.150	0.147	0.034	0.150	0.031	0.128	0.030	0.030	0.030	0.150	0.030	0.030	0.030	0.030
DNK	0.030	0.030	0.094	0.030	0.037	0.035	0.030	0.148	0.030	0.150	0.030	0.150	0.030	0.057
ESP	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.030	0.150	0.150	0.150	0.030	0.150	0.031
EST	0.148	0.078	0.030	0.030	0.030	0.150	0.150	0.150	0.084	0.030	0.030	0.030	0.030	0.030
FIN	0.150	0.150	0.030	0.124	0.044	0.150	0.030	0.127	0.030	0.030	0.030	0.030	0.030	0.046
FRA	0.134	0.150	0.146	0.030	0.150	0.030	0.030	0.030	0.030	0.030	0.150	0.030	0.030	0.030
GBR	0.030	0.032	0.030	0.095	0.030	0.141	0.066	0.150	0.036	0.030	0.030	0.030	0.150	0.150
GRC	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.150	0.130	0.030	0.030	0.030
HUN	0.040	0.030	0.120	0.030	0.150	0.030	0.030	0.030	0.030	0.150	0.150	0.030	0.150	0.030
IRL	0.030	0.150	0.030	0.030	0.148	0.150	0.043	0.030	0.067	0.030	0.030	0.030	0.119	0.113
ISL	0.030	0.030	0.030	0.030	0.030	0.030	0.133	0.150	0.150	0.030	0.142	0.149	0.036	0.030
ISR	0.030	0.030	0.030	0.136	0.030	0.030	0.038	0.030	0.030	0.142	0.144	0.150	0.150	0.030
ITA	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.128	0.149	0.030	0.030	0.032
JPN	0.138	0.030	0.030	0.072	0.030	0.150	0.149	0.094	0.036	0.030	0.150	0.030	0.030	0.030
KOR	0.030	0.030	0.030	0.150	0.030	0.148	0.147	0.135	0.030	0.150	0.030	0.030	0.030	0.030
LTU	0.030	0.030	0.030	0.030	0.077	0.030	0.150	0.030	0.133	0.030	0.100	0.150	0.030	0.150
LUX	0.030	0.030	0.141	0.117	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.053	0.150	0.030
LVA	0.030	0.030	0.128	0.030	0.110	0.030	0.150	0.150	0.150	0.030	0.030	0.030	0.030	0.072
MEX	0.030	0.030	0.030	0.030	0.030	0.030	0.127	0.030	0.150	0.030	0.033	0.150	0.150	0.150

Table 4 (continued)

	SCI	MATH	READ	EXCE	INCL	RESI	INCS	INCM	INCR	BELO	BULL	GOAL	TEACH	VALU
NLD	0.030	0.030	0.150	0.150	0.030	0.129	0.030	0.031	0.030	0.150	0.150	0.030	0.030	0.030
NOR	0.030	0.030	0.030	0.030	0.030	0.030	0.130	0.150	0.150	0.150	0.030	0.150	0.030	0.030
NZL	0.149	0.147	0.030	0.150	0.030	0.032	0.030	0.030	0.030	0.030	0.030	0.031	0.131	0.150
POL	0.150	0.150	0.143	0.150	0.076	0.090	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030
PRT	0.030	0.032	0.030	0.030	0.127	0.030	0.030	0.030	0.030	0.150	0.150	0.031	0.150	0.150
SVK	0.038	0.030	0.150	0.150	0.149	0.030	0.030	0.030	0.030	0.030	0.150	0.030	0.030	0.122
SVN	0.150	0.030	0.149	0.030	0.132	0.150	0.030	0.030	0.030	0.030	0.149	0.030	0.030	0.030
SWE	0.030	0.150	0.033	0.150	0.034	0.112	0.030	0.070	0.030	0.030	0.150	0.030	0.036	0.116
TUR	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.134	0.150	0.030	0.146	0.030	0.030	0.150
USA	0.136	0.150	0.030	0.031	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.142

Table 5 Normalised weights ω^- for the global data

	SCI	MATH	READ	EXCE	INCL	RESI	INCS	INCM	INCR	BELO	BULL	GOAL	TEACH	VALU
AUS	0.030	0.030	0.150	0.030	0.139	0.040	0.030	0.030	0.030	0.150	0.150	0.131	0.030	0.030
AUT	0.030	0.077	0.030	0.030	0.030	0.035	0.150	0.150	0.133	0.030	0.030	0.034	0.091	0.150
BEL	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.030	0.030	0.030	0.130	0.150
CAN	0.030	0.030	0.145	0.030	0.135	0.030	0.030	0.030	0.030	0.150	0.150	0.030	0.030	0.150
CHE	0.030	0.036	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.030	0.030	0.120	0.033	0.150
CHL	0.130	0.030	0.150	0.150	0.150	0.150	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030
COL	0.150	0.150	0.150	0.130	0.150	0.030	0.030	0.150	0.030	0.150	0.030	0.030	0.030	0.030
CZE	0.030	0.030	0.030	0.030	0.030	0.030	0.143	0.150	0.137	0.150	0.030	0.030	0.150	0.030
DEU	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.030	0.044	0.106	0.040	0.150
DNK	0.150	0.035	0.030	0.150	0.030	0.031	0.120	0.030	0.030	0.030	0.150	0.030	0.150	0.034
ESP	0.150	0.081	0.150	0.150	0.030	0.030	0.030	0.079	0.030	0.030	0.030	0.150	0.030	0.030
EST	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.137	0.150	0.150	0.143
FIN	0.030	0.030	0.098	0.030	0.030	0.030	0.030	0.030	0.062	0.150	0.150	0.150	0.150	0.030
FRA	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.030	0.030	0.150	0.030	0.130
GBR	0.030	0.030	0.144	0.030	0.136	0.030	0.031	0.030	0.030	0.150	0.150	0.150	0.030	0.030
GRC	0.150	0.069	0.093	0.150	0.030	0.150	0.030	0.030	0.030	0.030	0.030	0.030	0.148	0.030
HUN	0.030	0.030	0.030	0.030	0.030	0.130	0.150	0.150	0.150	0.030	0.030	0.150	0.030	0.030
IRL	0.119	0.030	0.040	0.150	0.030	0.030	0.030	0.031	0.030	0.150	0.150	0.150	0.030	0.030
ISL	0.150	0.150	0.030	0.150	0.130	0.150	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030
ISR	0.150	0.055	0.148	0.030	0.150	0.150	0.030	0.107	0.030	0.030	0.030	0.030	0.030	0.030
ITA	0.150	0.132	0.030	0.150	0.030	0.148	0.030	0.030	0.030	0.030	0.030	0.150	0.030	0.030
JPN	0.030	0.150	0.030	0.030	0.079	0.030	0.030	0.030	0.030	0.150	0.030	0.150	0.081	0.150
KOR	0.030	0.030	0.030	0.030	0.082	0.030	0.030	0.030	0.087	0.030	0.150	0.150	0.150	0.141
LTU	0.030	0.078	0.030	0.150	0.030	0.148	0.030	0.085	0.030	0.150	0.030	0.030	0.150	0.030
LUX	0.030	0.030	0.030	0.030	0.030	0.140	0.150	0.150	0.150	0.030	0.030	0.030	0.030	0.030
LVA	0.030	0.030	0.030	0.130	0.030	0.030	0.030	0.030	0.030	0.150	0.150	0.150	0.150	0.140
MEX	0.150	0.150	0.150	0.150	0.130	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030

Table 5 (continued)

	SCI	MATH	READ	EXCE	INCL	RESI	INCS	INCM	INCR	BELO	BULL	GOAL	TEACH	VALU
NLD	0.030	0.150	0.030	0.030	0.032	0.030	0.150	0.030	0.128	0.030	0.030	0.150	0.030	0.150
NOR	0.150	0.030	0.127	0.035	0.148	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.150	0.150
NZL	0.030	0.030	0.053	0.030	0.030	0.030	0.150	0.150	0.106	0.150	0.150	0.031	0.030	0.030
POL	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.127	0.150	0.150	0.033	0.150	0.150
PRT	0.030	0.030	0.030	0.055	0.030	0.123	0.150	0.150	0.150	0.030	0.030	0.132	0.030	0.030
SVK	0.030	0.030	0.030	0.030	0.030	0.030	0.034	0.150	0.143	0.142	0.030	0.150	0.141	0.030
SVN	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.131	0.150	0.030	0.150	0.149	0.150
SWE	0.049	0.030	0.058	0.030	0.031	0.030	0.150	0.040	0.149	0.142	0.030	0.150	0.079	0.032
TUR	0.061	0.150	0.150	0.150	0.100	0.033	0.030	0.030	0.030	0.145	0.030	0.030	0.031	0.030
USA	0.030	0.030	0.150	0.030	0.030	0.115	0.030	0.045	0.150	0.150	0.150	0.030	0.030	0.030

Table 6 Composite indicator in terms of separate dimensions

	Total			Academic			Equity			Students		
	Global	Male	Female	Global	Male	Female	Global	Male	Female	Global	Male	Female
AUS	1.003	0.998	1.005	1.017	1.015	1.021	1.012	1.008	1.014	0.974	0.975	0.970
AUT	0.977	0.990	0.978	1.005	1.017	1.006	0.996	0.996	0.995	0.977	0.978	0.977
BEL	1.001	1.004	1.002	1.027	1.039	1.027	0.979	0.980	0.978	0.995	0.986	0.997
CAN	1.094	1.089	1.087	1.060	1.040	1.061	1.052	1.059	1.038	0.982	0.990	0.987
CHE	1.024	1.026	1.024	1.032	1.032	1.031	0.990	0.988	0.990	1.003	1.007	1.004
CHL	0.757	0.772	0.759	0.819	0.827	0.820	0.870	0.884	0.871	1.063	1.056	1.063
COL	0.712	0.744	0.705	0.730	0.751	0.728	0.935	0.949	0.935	1.044	1.044	1.037
CZE	0.927	0.910	0.929	1.014	1.009	1.014	0.983	0.991	0.935	0.930	0.909	0.933
DEU	1.019	1.019	1.019	1.027	1.030	1.025	0.995	0.990	0.995	0.998	0.999	0.999
DNK	1.077	1.090	1.078	1.009	1.007	1.010	1.013	1.015	1.012	1.053	1.066	1.054
ESP	1.013	1.036	1.017	0.953	0.967	0.953	1.006	1.015	1.009	1.056	1.056	1.057
EST	1.032	1.030	1.052	1.031	1.029	1.054	1.075	1.079	1.072	0.930	0.928	0.931
FIN	1.089	1.066	1.093	1.044	1.027	1.048	1.043	1.027	1.043	1.000	1.011	1.000
FRA	0.946	0.956	0.946	1.008	1.011	1.008	0.976	0.980	0.977	0.962	0.965	0.961
GBR	1.039	1.040	1.020	1.025	1.018	1.018	1.027	1.026	1.015	0.988	0.996	0.987
GRC	0.822	0.802	0.821	0.901	0.885	0.900	0.942	0.931	0.942	0.968	0.974	0.968
HUN	0.918	0.942	0.918	0.985	0.988	0.985	0.940	0.952	0.940	0.991	1.001	0.991
IRL	1.025	1.025	1.020	1.020	1.019	1.019	1.028	1.035	1.028	0.978	0.972	0.975
ISL	1.035	1.008	1.032	0.981	0.970	0.980	1.004	0.992	1.004	1.050	1.047	1.049
ISR	0.943	0.930	0.944	0.971	0.959	0.973	0.943	0.941	0.942	1.030	1.030	1.030
ITA	0.955	0.972	0.954	0.971	0.983	0.971	0.994	1.003	0.993	0.990	0.986	0.989
JPN	1.103	1.129	1.094	1.043	1.068	1.025	1.054	1.045	1.058	1.003	1.011	1.008
KOR	1.174	1.200	1.175	1.064	1.075	1.067	1.054	1.050	1.051	1.047	1.063	1.048
LTU	0.918	0.896	0.920	0.967	0.966	0.966	0.980	0.978	0.980	0.968	0.949	0.972
LUX	0.921	0.915	0.921	0.989	0.988	0.990	0.942	0.938	0.943	0.989	0.987	0.987

Table 6 (continued)

	Total			Academic			Equity			Students		
	Global	Male	Female	Global	Male	Female	Global	Male	Female	Global	Male	Female
LVA	0.859	0.842	0.860	0.978	0.979	0.978	1.011	1.009	1.010	0.869	0.852	0.870
MEX	0.740	0.754	0.736	0.718	0.730	0.714	0.947	0.952	0.947	1.087	1.085	1.087
NLD	1.025	1.027	1.025	1.039	1.041	1.039	1.020	1.009	1.021	0.967	0.978	0.967
NOR	1.081	1.062	1.082	1.013	1.005	1.014	1.019	1.018	1.018	1.048	1.037	1.049
NZL	0.998	0.998	0.999	1.034	1.035	1.034	1.003	1.003	1.004	0.963	0.962	0.961
POL	0.950	0.930	0.956	1.050	1.052	1.048	1.015	1.007	1.022	0.892	0.878	0.892
PRT	1.091	1.103	1.090	1.003	1.007	1.002	0.990	0.996	0.989	1.098	1.100	1.100
SVK	0.824	0.813	0.839	0.968	0.968	0.966	0.968	0.970	0.964	0.880	0.866	0.901
SVN	0.914	0.916	0.917	1.019	1.015	1.020	1.022	1.014	1.022	0.878	0.890	0.879
SWE	1.071	1.081	1.065	1.028	1.022	1.030	1.015	1.014	1.010	1.027	1.043	1.023
TUR	0.903	0.871	0.894	0.905	0.911	0.903	0.995	0.993	0.979	1.003	0.963	1.011
USA	1.019	1.031	1.020	1.007	1.008	1.010	0.996	0.999	0.996	1.016	1.023	1.015

Table 7 Comparative results of the tripartite decomposition

	PC_i	OWN_i	BP_i	$W*_i$
AUS	**1.007**	**1.058**	0.959	0.993
AUT	0.988	**1.002**	0.967	**1.020**
BEL	0.998	**1.027**	0.964	**1.008**
CAN	0.998	**1.021**	0.969	**1.008**
CHE	0.999	**1.021**	0.977	**1.001**
CHL	0.983	**1.007**	0.947	**1.032**
COL	0.948	0.945	0.964	**1.041**
CZE	**1.022**	**1.038**	0.971	**1.014**
DEU	**1.000**	**1.019**	0.972	**1.010**
DNK	0.989	**1.028**	0.956	**1.007**
ESP	0.981	**1.006**	0.952	**1.024**
EST	**1.021**	**1.055**	0.952	**1.016**
FIN	**1.025**	**1.110**	0.951	0.972
FRA	0.989	**1.021**	0.962	**1.008**
GBR	0.981	**1.032**	0.952	0.999
GRC	**1.024**	**1.086**	0.968	0.974
HUN	0.975	1.000	0.951	**1.025**
IRL	0.996	**1.049**	0.949	**1.001**
ISL	**1.024**	**1.089**	0.963	0.977
ISR	1.015	**1.079**	0.946	0.994
ITA	0.982	**1.010**	0.961	**1.012**
JPN	0.969	**1.037**	0.952	0.982
KOR	0.980	**1.004**	0.954	**1.024**
LTU	**1.026**	**1.100**	0.953	0.979
LUX	**1.007**	**1.056**	0.961	0.993
LVA	**1.021**	**1.084**	0.955	0.985
MEX	0.976	**1.004**	0.959	**1.014**
NLD	0.999	**1.030**	0.955	**1.015**
NOR	**1.019**	**1.076**	0.962	0.985
NZL	**1.001**	**1.033**	0.966	**1.003**
POL	**1.028**	**1.086**	0.953	0.994
PRT	0.988	**1.007**	0.958	**1.025**
SVK	**1.032**	**1.062**	0.954	**1.019**
SVN	**1.001**	**1.049**	0.947	**1.007**
SWE	0.985	**1.014**	0.961	**1.012**
TUR	**1.027**	**1.086**	0.972	0.972
USA	0.990	**1.031**	0.949	**1.011**

Values printed in bold type indicate a positive contribution

Author contributions IC contributes to the development of the methodology for the construction of the composite indicators. MMS computed and analysed the main results for the case study. Both authors contribute to the discussion of the results and elaboration of main conclusions.

Data Availability The datasets analysed during the current study are available from the PISA 2018 report repository, see at https://www.oecd.org/pisa/data/2018database/

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Agasisti, T., Avvisati, F., Borgonovi, F., Longobardi, S. (2018). Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA. OECD Education Working Papers 167, 1–40. https://doi.org/10.1787/e22490ac-en.

Arcos, E. G., Figueroa, V. A., Miranda, C., & Ramos, C. (2007). State of the art and foundations for the gender indicators construction in education. *Estudios Pedagogicos, 2*, 121–130.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.

Bas, M. C., & Carot, J. M. (2020). A model for developing an academic activity index for higher education instructors based on composite indicators. *Educational Policy*. https://doi.org/10.1177/0895904820951123

Benito, M., & Romera, R. (2011). Improving quality assessment of composite indicators in university rankings: A case study of French and German universities of excellence. *Scientometrics, 89*(1), 153–176.

Bericat, E. (2012). The European gender equality index: Conceptual and analytical issues. *Social Indicators Research, 108*(1), 1–28.

Bericat, E., & Sanchez-Bermejo, E. (2016). Structural gender equality in Europe and its evolution over the first decade of the twentyfirst century. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, 127*(1–3), 55–81.

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research, 87*(2), 425–469.

Blancas, F. J., Contreras, I., & Ramirez-Hurtado, J. M. (2012). Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives. *Journal of the Operational Research Society, 64*(5), 668–678.

Blossfeld, P. N., Blossfeld, G. J., & Blossfeld, H. P. (2016). Changes in Educational Inequality in Cross-National Perspective. In M. J. Shanahan, J. T. Mortimer, & M. K. Johnson (Eds.), *Handbook of the Life Course* (pp. 223–247). Cham: Springer. https://doi.org/10.1007/978-3-319-20880-0_10

Casas, F., Bello, A., González, M., & Aligué, M. (2013). Children's subjective well-being measured using a composite index: What Impacts Spanish first-year secondary education students' subjective well-being? *Child Indicators Research, 6*, 433–460.

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics, 106*(1), 1–25.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.

Cherchye, L., Moesen, W., Puyenbroeck, T. (2003). Legitimately Diverse, yet comparable: On synthesising social inclusion performance in the EU. CES Discussion paper series 03.01, KU Leuven.

Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. (2007). An Introduction to 'Benefit of the Doubt' composite indicators. *Social Indicators Research, 82*(1), 111–145.

Contreras, I., & Hinojosa, M. A. (2020). A note on "The cross-efficiency in the optimistic–pessimistic framework." *Operational Research International Journal, in Press.* https://doi.org/10.1007/s12351-019-00484-2

Docampo, D., & Cram, L. (2017). Academic performance and institutional resources: A cross-country analysis of research universities. *Scientometrics, 110*(2), 739–764.

Domínguez, C., Segovia-González, M. M., & Contreras, I. (2021). A multiplicative composite indicator to evaluate educational systems in OECD countries. *Compare: A Journal of Comparative and International Education.* https://doi.org/10.1080/03057925.2020.1865791

Downey, D. B., & Condron, D. J. (2016). Fifty years since the Coleman report: Rethinking the relationship between schools and inequality. *Sociology of Education, 89*(3), 207–220. https://doi.org/10.1177/0038040716651676

Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: A social choice approach. *Journal of Environmental Economic Management, 47*(2), 270–283.

El Gibari, S., Gómez, T., & Ruiz, F. (2018). Evaluating university performance using reference point based composite indicators. *Journal of Informetrics, 12*(4), 1235–1250.

Esty, D. C., Levy, M., Srebotnjak, T., & Sherbinin, A. (2005). *Environmental sustainability index: Benchmarking national environmental stewardship*. New Haven: Yale Center of Environmental Law and Policy.

Ferrant, G. (2014). The multidimensional gender inequalities index (MGII): A descriptive analysis of gender inequalities using MCA. *Social Indicators Research, 115*(2), 653–690.

Ferrara, A. R., & Nisticò, R. (2013). Well-being indicators and convergence across italian regions. *Applied Research in Quality of Life, 8*(1), 15–44.

Giambona, F., & Vassallo, E. (2014). Composite indicator of social inclusion for European countries. *Social Indicators Research, 116*(1), 269–293.

Gnaldi, M., & Ranalli, M. G. (2016). Measuring university performance by means of composite indicators: A robustness analysis of the composite measure used for the benchmark of Italian universities. *Social Indicators Research, 129*(2), 659–675.

Goodenow, C., & Grady, K. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *The Journal of Experimental Education, 62*(1), 60–71.

Hadjar, A., & Buchmann, C. (2016). Education systems and gender inequalities in educational attainment. In A. Hadjar & C. Gross (Eds.), *Education Systems and Inequalities: International Comparisons.* Bristol University Press.

Jovanovic-Gavrilovic, B., & Radivojevic, B. (2017). Education of population for the future and the future of education. *Stanovnistvo, 55*(1), 63–85.

Khodabakhshi, M., & Aryavash, K. (2012). Ranking all units in data envelopment analysis. *Applied Mathematics Letters, 25*(12), 2066–2070.

Liu, J., Peng, P., & Luo, L. (2019). The relation between family socioeconomic status and academic achievement in China: A meta-analysis. *Educational Psychology Review, 32*(2), 1–28.

Ma, L., Luo, H., & Xiao, L. (2021). Perceived teacher support, self-concept, enjoyment and achievement in reading: A multilevel mediation model based on PISA 2018. *Learning and Individual Differences, 85*(101947), 1–9.

Mancebon, M. J., & Bandres, E. (1999). Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results. *Education Economics, 7*(2), 131–152.

Mikk, J., Krips, H., Säälik, U., & Kalk, K. (2016). Relationships between student¨ perception of teacher-student relations and PISA results in mathematics and science. *International Journal of Science and Mathematics Education, 14*, 1437–1454.

Mitrovic, M., Markovic, M., & Zdravkovic, S. (2016). Statistical approach for ranking OECD countries based on composite GICSES index and I-distance method. *Emerging Trends in the Development and Application of Composite Indicators* (pp. 324–348). Pennsylvania: IG-Blobal.

Monte, A., & Schoier, G. (2022). A Multivariate Statistical Analysis of Equitable and Sustainable Well-Being Over Time. *Social Indicators Research, 161*(2–3), 735–750.

Morris, A. (2011). Student standardised testing: Current practices in OECD countries and a literature review. Education Working Paper 65.

Moser, A. (2007). Gender and Indicators. *Bridge Development-Gender*. UK: Institute of Development Studies.

Murias, P., de Miguel, C., & Rodríguez, D. (2007). A Composite Indicator for University quality assessment: The case of Spanish higher education system. *Social Indicators Research, 89*(1), 129–146.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E. (2005a). Handbook on constructing composite indicators: Methodology and user guide. OECD Statistics Working Papers.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for composite indicator building*. European Commission: Institute for the Protection and Security of the Citizen.

OECD. (2008). *Handbook on constructing composite indicators*. Paris: OECD Publishing.

OECD. (2017). *PISA 2015 Assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving*. Paris: OECD Publishing. https://doi.org/10.1787/9789264281820-en

OECD. (2021). *PISA 2018 Database*. Paris: OECD Publishing.

OECD. (2021b). *Education at a Glance 2021: OECD Indicators*. OECD Publishing.

Rees, G., & Main, G. (2015). *Children's views on their lives and well-being in 15 countries: A report on the children's worlds survey, 2013–14*. York, UK: Children's Worlds Project.

Roemer, J. E., & Trannoy, A. (2015). Equality of opportunity. *Handbook of Income Distribution, 2*, 217–300. https://doi.org/10.1016/B978-0-444-594280.00005-9

Rothstein, R. (2000). Toward a composite index of school performance. *The Elementary School Journal, 100*(5), 409–441.

Saisana, M., & Tarantola, S. (2002). *State-of-the-Art report on current methodologies and practices for composite-indicator development*. European Commission: Joint Research Centre.

Selvitopu, A., & Kaya, M. (2021). A meta-analytic review of the effect of socioeconomic status on academic performance. *Journal of Education*. https://doi.org/10.1177/00220574211031978

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Sirin, S. (2005). Socioeconomic status and academic achievement: A metaanalytic review of research. *Review of Educational Research, 75*(3), 417453.

Stumbriene, D., Camanho, A. S., & Jakaitiene, A. (2020). The performance of education systems in the light of Europe 2020 strategy. *Annals of Operations Research, 288*, 577–608.

Torres-Salinas, D., Moreno-Torres, J. G., Delgado-López-Cózar, E., & Herrera, F. (2011). A methodology for Institution-Field ranking based on a bidimensional analysis: The IFQ2A index. *Scientometrics, 88*, 771–786.

UNESCO (2015) Incheon Declaration. Education 2030: Towards inclusive and equitable quality education and lifelong learning for all. http://unesdoc.unesco.org/images/0023/002338/233813M.pdf.

United Nations. (2020). *The sustainable development goals report 2020*. United Nations: UN DESA.

Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with Benefit-of-the Doubt weights. *European Journal of the Operational Research, 256*(3), 1004–1014.

Verbunt, P., & Rogge, N. (2018). Geometric composite indicators with compromise Benefit-of-the Doubt weights. *European Journal of the Operational Research, 264*(1), 388–401.

Villar, A. (2017). Performance, inclusion and excellence: An index of educational achievements for PISA. *Advances in Social Sciences Research Journal, 4*(3), 110–115.

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461–481.

Yoon, K. P., & Hwang, C. L. (1995). *Multiple-attribute decision-making: An introduction*. Sage Publications.

Zhou, P., Ang, B. W., & Poh, K. L. (2006). Comparing aggregation methods for constructing the composite environmental index: An objective measure. *Ecological Economics, 59*(3), 305–311.

Zhou, P., Ang, B. W., & Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research, 96*(1), 169–181.