



# Learning mobility in European higher education: How has the Union's flagship initiative progressed?

Miguel Alves Pereira<sup>1</sup> · Giovanna D'Inverno<sup>2,3</sup> · Ana Santos Camanho<sup>4</sup>

Received: 19 December 2023 / Accepted: 26 July 2024  
© The Author(s) 2024

## Abstract

In 2010, the European Commission set out the development of an economy based on knowledge and innovation as one of the priorities of its Europe 2020 strategy for smart, sustainable, and inclusive growth. This culminated in the 'Youth on the Move' flagship initiative, aimed at enhancing the performance and international attractiveness of Europe's higher education institutions and raising the Union's overall education and training levels. Therefore, it is relevant to assess the performance of the 'Youth on the Move' initiative via the creation of composite indicators (CIs) and, ultimately, monitor the progress made by European countries in creating a positive environment supporting learner mobility. For this reason, we make use of the CI-building 'Benefit-of-the-Doubt' approach, in its robust and conditional setting to account for outliers and the human development of those nations, to exploit the European Commission's Mobility Scoreboard framework between 2015/2016 and 2022/2023. Furthermore, we incorporate the value judgements of experts in the sector to construct utility scales and compute weight restrictions through multi-criteria decision analysis. This enables the conversion of ordinal scales into interval ones based on knowledgeable information about reality in higher education. In the end, the results point to a slight performance improvement, but highlight the need to improve the 'Recognition of learning outcomes', 'Foreign language preparation', and 'Information and guidance'.

**Keywords** Benefit-of-the-Doubt · Composite indicators · Utility scales · European education systems · Learning mobility

---

✉ Miguel Alves Pereira  
miguelalvespereira@tecnico.ulisboa.pt

Giovanna D'Inverno  
giovanna.dinverno@unipi.it

Ana Santos Camanho  
acamanho@fe.up.pt

<sup>1</sup> CEGIST, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, Lisboa 1049-001, Portugal

<sup>2</sup> Dipartimento di Economia e Management, Università di Pisa, Via Cosimo Ridolfi 10, Pisa 56124, Italia

<sup>3</sup> Faculty of Economics and Business, KU Leuven, Naamsestraat 69, Leuven 3000, Belgium

<sup>4</sup> Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, Porto 4200-465, Portugal

# 1 Introduction

Ambition has always driven “the old continent” to explore new horizons. From the Renaissance to the Enlightenment and colonialism and imperialism to the Industrial Revolution, Europe has historically been the main driving force of the (Western) world. One of the continent’s most recent and progressive endeavours is its 10-year Europe 2020 strategy. Proposed by the European Commission in 2010, Europe 2020 aimed at marking a new beginning in the wake of the 2008 economic and financial crisis via “smart, sustainable and inclusive growth” (European Commission, 2010). The strategy identified five key targets that the European Union (EU) should work towards, namely in terms of employment, research and innovation, climate change and energy, education, and combating poverty. For this reason, seven flagship initiatives were created, whose implementation relied on framework programmes such as the well-known 6-year Horizon 2020 and its 80 billion Euro budget.

After the proposal put forward by the European Commission (2010), the Council of the European Union (2011) recommended launching the ‘Youth on the Move’ flagship initiative for the sake of “promoting the learning mobility of young people”. According to the same source, learning mobility implies moving beyond one’s country’s borders to acquire new knowledge, skills, and competencies. It is seen as a cornerstone for strengthening the employability of young people, as well as increasing their awareness of other cultures, fostering their personal growth, stimulating their creativity, and encouraging their active citizenship. Indeed, the Council of the European Union believes that such mobility at the higher education level not only increases the chances of professional mobility after graduation but also nurtures the openness, communality, accessibility, and efficiency of education and training systems conducive to a knowledge-intensive society.

Despite a majority of positive reactions to Europe 2020, scepticism about the strategy prevailed in the sense that doubts were raised about the five key targets being the right priorities, at first, and their actual success at the end of the implementation period, at last. Nevertheless, one of these targets seems to stand out: education. As a matter of fact, the European Commission (2010) stated that “better education levels help employability”, which itself “helps to reduce poverty”. Besides, the links between (higher) education and research and innovation are clear, not to mention the latter’s contribution to “fighting climate change and create [*sic*] new business and employment opportunities” (European Commission, 2010).

Through the ‘Youth on the Move’ flagship initiative, EU countries should seek to not only improve the performance of their higher education institutions but also increase the overall quality of all their education and training levels. This could be achieved by combining excellence and quality, promoting mobility, and improving the employability of young people. Particularly, the European Commission stated that its first plan of action at the EU level would be to focus on integrating and enhancing EU mobility, university, and researchers’ programmes (e.g., Erasmus, Erasmus Mundus, TEMPUS,<sup>1</sup> Marie Skłodowska-Curie Actions) with national programmes and resources to provide flexible access to continuous learning beyond traditional boundaries.

Therefore, evaluating the ‘Youth on the Move’ flagship initiative across Europe is vital for tracking the evolution of EU countries and their European peers—such as those belonging to the European Economic Area (EEA) and the European Free Trade Association (EFTA)—in terms of learning mobility and providing insights with education policy-making relevance. Accordingly, based on the European Commission’s Mobility Scoreboard regarding higher education, which rests on information provided by the Eurydice Network, we resort to the

<sup>1</sup> Since January 1, 2014, the TEMPUS program was rebranded as the Erasmus+ program.

popular non-parametric frontier ‘Benefit-of-the-Doubt’ (BoD) approach to assess the performance of 36 countries and macro-regions in the network’s reference years of 2015/2016, 2018/2019, and 2022/2023 in order to create a yearly composite indicator (CI). Our model is extended to a robust and conditional setting since it reduces the influence of outlying data and takes the countries’ contextual environment into account. Additionally, we adopt the trending multi-criteria decision analysis (MCDA) ‘Deck of Cards’ method to convert Eurydice’s qualitative data into quantitative data and incorporate the value judgements of experts in the education sector concomitantly, on the one hand, and create more realistic weight restrictions, on the other hand. This way we generate preference-incorporated interval scales and relative importance multipliers. Furthermore, we study the performance evolution of this education policy by means of the global Malmquist productivity index (*MPI*) to employ a metafrontier that overcomes circularity issues regarding the simultaneous analysis of three periods. Ultimately, our study contributes to the literature both theoretical- (original hybrid BoD and MCDA approach, given the combination of a robust conditional setting, the use of weight restrictions, and the incorporation of value judgements via the ‘Deck of Cards’ method) and empirically (unprecedented assessment of learning mobility performance using more recent and broader data from an official database).

The remaining sections of the paper are arranged as follows. In Sect. 2, a review of the literature is provided on the usage of CIs created using the BoD approach, giving special attention to the education sector. The knowledge gap and contributions of the paper to the existing literature are also highlighted. In Sect. 3, we describe the learning mobility performance assessment framework, whose results are discussed in Sect. 4. At last, Sect. 5 concludes our work and identifies areas for future research.

## 2 Knowledge gap

Depending on the EU nation, education strategy is developed at the local, regional, or national level to ensure alignment with the policies put forward by the European Commission. Therefore, the performance of education systems should be closely monitored, for the sake of securing the EU’s leading role in the global education market (Camanho, Stumbriene, Barbosa, and Jakaitiene, 2023). This requires gathering data regarding a set of outcome indicators and using frontier techniques to provide a big picture of the efficiency and effectiveness of education systems.

### 2.1 Measuring efficiency

De Witte and López-Torres (2017) conducted a review of the literature on efficiency in education and showed that most publications are concerned with higher education and the efficient use of resources, i.e., the capacity of institutions to reduce expenses while maintaining the same student performance or research production, for instance. Overall, regardless of the level of education being analysed (which ranges from primary and secondary education to tertiary education), the literature addresses a wide variety of levels of analysis, namely microscopic (e.g., students, classrooms, departments), regional (e.g., school districts, macro-regions), and international (e.g., countries’ education systems). Besides, the authors mention that frontier techniques, especially non-parametric ones of which Data Envelopment Analysis (DEA) is at the forefront, are among the most commonly employed in the field.

Given the far-reaching scope of the ‘Youth on the Move’ flagship initiative, it may be wise to direct our focus to international benchmarking comparisons, specifically those that used frontier techniques to carry out cross-country performance assessments. First, a survey of the literature indicates that only a handful of studies is about tertiary education (Agasisti, 2011; Bogetoft, Heinesen, and Tranæs, 2015; Ahec Sonje, Deskar-Skrbic, and Sonje, 2018; Stumbriene, Camanho, and Jakaitiene, 2020), although only Agasisti (2011) focuses exclusively on this stage. Second, the vast majority of papers use data from large-scale comparable international datasets, namely the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). Third, in line with the findings of De Witte and López-Torres (2017), DEA is, indeed, the most used education efficiency measurement technique, at least at the country level (see, e.g., Agasisti et al. 2019; Giambona et al. 2011; Giménez et al. 2007). As for the evaluation of efficiency changes over time, the literature is scarce, with Agasisti (2014) having used the Malmquist index, Giménez et al. (2017) having used the global Malmquist index, and Giménez et al. (2019) having used the Malmquist-Luenberger index. Nevertheless, these publications only considered periods until 2012 and up to 29 European countries—always using DEA. Finally, only a few studies have introduced the use of conditional robust partial frontier methods to account for the role of the operating context on the production process (e.g., Cordero et al. 2018, 2017; De Witte & Kortelainen, 2013) or other recent extensions developed for the Malmquist productivity indices (e.g., Aparicio et al. 2021, 2022; Arbona et al. 2022). In these studies, efficiency and productivity are measured over the period spanning from 2006 to 2018.

## 2.2 Assessing performance

Alternatively, one can look at the efficiency of education systems from another standpoint: that of performance assessment. When assessing performance, the onus lies on the sole evaluation of outputs, i.e., resources do not constrain the problem. For this reason, the DEA-based BoD approach—developed by Cherchye et al. (2007)—must be used. Hence, we are essentially combining multiple outcomes into a single measure that eases the interpretation and communication of information, enables the monitoring of those indicators, and can be used to guide decision-making (Pereira, Camanho, Figueira, and Marques, 2021). These so-called CIs are, indeed, performance aggregation tools that tend to resort to DEA and DEA-based methods, which include the popular BoD approach, due to their appeal in terms of data-driven flexible weight optimisation (Nardo et al., 2008).

Some of the earliest works introducing the BoD approach to create a composite measure in education are by De Witte and Rogge (2010, 2011) and Rogge (2011). These studies evaluated the academic performance of teachers and faculty members in Belgium, considering both teaching and/or research performance. The first two papers introduced the robust and conditional approach to include the background conditions in the performance assessment. De Witte et al. (2013) studied the presence of economies of scope in the teaching-research nexus and found that, for instance, specialisation in teaching and research correlates with better academic performance. Karagiannis and Paschalidou (2017) examined research effectiveness by comparing the BoD approach with the model of Kao and Hung (2003). The authors found that there is a higher variability in the effectiveness scores estimated among alternative weighting schemes within each model than between models for particular weighing schemes. De Witte and Schiltz (2018) used the robust and conditional iteration of the BoD approach to measure and explain the organisational effectiveness of school districts. With this modelling choice, the authors intended to reduce the influence of outliers and account

for the exogenous environment. Szuwarzyński (2019) assessed the performance of public universities in Australia by means of the BoD approach with weight restrictions. Taking into account teaching and research key performance indicators, the author provided directions for future improvement for the worst-performing institutions. Silva et al. (2020) used the results of Portuguese students on national exams to assess the performance of secondary schools and benchmark them on their success in preparing students for success in higher education. The CI built by these authors was attained via a directional distance function BoD that intended to adopt a directional vector corresponding to the range of possible improvement for each school. Bas and Carot (2022) took advantage of the BoD approach's weight flexibility to assess teacher performance regarding several dimensions of academic activities. Afterwards, the authors classified them into clusters according to those dimensions. Szuwarzyński (2022) evaluated the efficiency of the graduation process in Australian public universities. Their "super-efficiency" BoD model with weight restrictions indicated that research-oriented universities achieve better results and overseas students perform better than domestic ones. Camanho et al. (2023) assessed the performance of European countries regarding the Strategic Framework for European Cooperation in Education and Training. Additionally, the authors studied the convergence of these countries in the 2009–2018 period, revealing that there is an improvement trend in most countries.

### 2.3 Overview

In the end, the knowledge gap is clear. First, this study is unprecedented since, as far as the authors are aware, there are no publications about learning mobility and the performance assessment of its inherent policies across European education systems, traditionally focused on broader characteristics. Second, it considers the most recently available data regarding the academic year 2022/2023. Third, it expands the usual sample size of performance assessments of European education systems to include data beyond EU countries, which entices interesting benchmarking possibilities. Fourth, it is the first to analyse this subject using data from the Eurydice Network database, offering a fresh perspective from the oft-used PISA and TIMSS databases. Lastly, the literature has yet to fully explore the potential of BoD approaches in education, especially those involving the robust conditional case (De Witte & Rogge, 2010; De Witte & Kortelainen, 2013; De Witte & Schiltz, 2018) and the application of MCDA (Agasisti, Munda, and Hippe, 2019). The simultaneous use of these techniques, especially when considering the 'Deck of Cards' method, is unique; furthermore, the inclusion of weight restrictions is rare (see, e.g., Szuwarzyński, 2019, 2022) and the use of MCDA to define them is a novelty.

## 3 Methodological framework

Despite the absence of a formal definition of a methodological framework within the academic community, it can generally be described "as a structured guide to completing a process or series of processes" (McMeekin, Wu, Germei, and Briggs, 2020). In fact, the authors acknowledge this lack of conventional guidance and propose a consensual approach to develop methodological frameworks, which we have adapted for our study. This approach involves three key phases: identifying data that inform the framework, developing the framework itself, and validating and refining it through iterative revisions. Specifically, McMeekin et al. (2020) highlight the integration of existing methods and guidelines, the incorporation

of expert experience, and the utilisation of empirical data to construct a robust framework. We employ this structured approach to systematically identify the relevant data inherent to our research problem (Sect. 3.1) and to develop the methods that operationalise this data (Sect. 3.2), ensuring our framework is both comprehensive and contextually appropriate.

### 3.1 Data and sample

The added value of learning mobility is materialised in an increase in opportunities (Nuzulismah, Azis, Sensuse, Kautsarina, and Suryono, 2021). However, the free movement of higher education students still faces several obstacles, such as the inflexible use of domestic grants to study abroad, the bureaucracy behind the recognition of qualifications, and the lack of straightforward access to information and guidance about mobility programs. Following the recommendation of the Council of the European Union (2011), the Eurydice—an information network funded by the European Commission that coordinates European and national efforts to provide policy-makers with credible and reliable information for education policy-making purposes—report produced by the European Education and Culture Executive Agency (2023) provides information on six Mobility Scoreboard key performance indicators in higher education. These are the indicators used in the proposed framework, which are part of an output vector  $Y_j^t = \{y_{1j}^t, \dots, y_{1r}^t, \dots, y_{1s}^t\}$ , with  $r = 1, \dots, s$  and described according to their latest iteration in Table 1.

Note that the Mobility Scoreboard is a framework that monitors the progress made by European countries under the ‘Youth on the Move’ flagship initiative. The Eurydice Network is the one that provides it with information<sup>2</sup> regarding the six thematic indicators specified by the Council of the European Union (2011). This qualitative information is collected through a survey completed by national experts and/or representatives of the Eurydice Network. Additionally, bear in mind that this network gathers qualitative data on top-level tertiary education policies and measures in three reference academic years—2015/2016, 2018/2019, and 2022/2023—covering 36 nations and macro-regions: all EU Member States, as well as EEA/EFTA countries and EU candidates (Bosnia and Herzegovina, Iceland, Liechtenstein, Montenegro, Norway, Serbia, and Türkiye). This removes any concerns about the “curse of dimensionality” that troubles DEA-based methods (Charles, Aparicio, and Zhu, 2019).

Due to missing data instances, Albania, Switzerland, North Macedonia, and the United Kingdom (comprised of England, Northern Ireland, Scotland, and Wales), were removed from the original data set. This action created a balanced panel data set. Furthermore, the Mobility Scoreboard disclaims that ‘Support to disadvantaged learners’ ( $y_4$ ) and ‘Recognition of learning outcomes’ ( $y_5$ ) are only comparable between 2018/2019 and 2022/2023 due to changes in their definitions.

### 3.2 Methods

The methods that comprise the process behind the proposed methodological framework consist of three steps. First, the ‘Deck of Cards’ method is detailed in Sect. 3.2.1. Second, the robust conditional ‘Benefit-of-the-Doubt’ (BoD) approach with weight restrictions is described in Sect. 3.2.2. Third, the global *MPI* (*GMPI*) is outlined in Sect. 3.2.3.

<sup>2</sup> Available at <https://national-policies.eacea.ec.europa.eu/mobility-scoreboard/higher-education/graphical-overview>.

**Table 1** Mobility Scoreboard six thematic indicators

Variable	Indicator	Description
y1	'Information and guidance'	It looks at the extent to which national/top-level authorities facilitate easy, universal access to information on outward learning mobility and support the delivery of high-quality, individualised guidance to all students throughout the education system. In particular, top-level information strategies, central online portals, the oversight of individualised services, and the participation of multipliers are examined
y2	'Foreign language preparation'	It revolves around the amount of time that a foreign language must be taught in full-time education, starting in preschool and continuing through upper secondary education. It looks at the length of time when (at least) one foreign language is required of all students, as well as the time when (at least) two foreign languages are required at the same time. It takes into account general and vocational education as well as educational tracks or courses that provide direct entry to higher education
y3	'Portability of grants and loans'	It focuses on the portability of domestic support, i.e., the ability of students to transfer their domestic public grant or loan to another European higher education system. When it comes to portable domestic support, it looks at whether portability applies to both short-term (credit mobility) and long-term (full-degree) study abroad. Further portability limits (such as national restrictions) are taken into consideration too
y4	'Support to disadvantaged learners'	It addresses how much mobility support specifically targets underprivileged students. It builds upon four main pillars of top-level support: financial support, top-level recommendations or incentives given to higher education institutions to implement targeted measures supporting the participation of disadvantaged students in mobility programmes, monitoring of mobility flows, quantitative policy objectives on the participation of disadvantaged learners in mobility programmes, and monitoring of mobility flows
y5	'Recognition of learning outcomes'	It measures development in the area of the European Credit Transfer and Accumulation System (ECTS), which supports international study mobility and has fundamental principles that must be effectively applied. It focuses particularly on the monitoring of the implementation of learning outcomes and student workload in ECTS by external quality assurance organisations, the use of ECTS supporting documents, prompt credit recognition and transfer, the existence of credit recognition appeals procedures, and statistical grade distribution tables
y6	'Recognition of qualifications'	It aims to evaluate national-level progress in modifying the practice of qualifications recognition to ensure the automatic recognition of qualifications from other European nations. On the one hand, it keeps track of whether qualifications are automatically recognised within the European Higher Education Area (EHEA) or whether additional procedures are in place for qualification recognition. On the other hand, it discusses some of the circumstances under which recognition procedures for the credentials granted in other EHEA nations might be implemented in education systems



### 3.2.1 Step 1: The ‘Deck of Cards’ method

Prior to conducting the performance assessment exercise using the BoD approach, we need to ensure that the indicators it operationalises have quantitative scales. In particular, to convert the ordinal scales provided by the Eurydice Network regarding the European Commission’s Mobility Scoreboard, there are a few alternatives that typically range from using data *as is* to statistical techniques. However, perhaps more uncommonly, MCDA can be considered since, despite its complexity, provides a more structured approach capable of handling the imprecise nature inherent to the assumptions of statistical techniques (e.g., ordinal scale transformation, Likert scale transformation).

MCDA finds its foundations in operational research. In essence, it seeks to structure, model, and offer solutions to decision problems surrounded by multiple criteria (Pereira, Machete, Ferreira, and Marques, 2020). Since criteria tend to be in conflict and an optimal solution is rarely found, the preferences of the decision-makers (DMs) are employed as a differentiating factor (Pereira & Marques, 2022a). Amongst its several disciplines, which include mathematical programming, fuzzy sets, and outranking methods, one finds multi-attribute utility theory (MAUT; Greco et al. 2016). MAUT is particularly useful for quantifying the relative attractiveness of a specific outcome.

Additionally, the literature has been showing over the last decade a trend towards the use of MCDA methods to construct composite indicators alongside the BoD approach (Gibari et al. 2019; Greco et al. 2019). However, in this particular context, the problem at hand required the use of an MCDA method to produce the performances of each country per indicator—not the final composite indicator, by simultaneously dealing with ordinal scales and incorporating the preference intensities of the DMs to consider attractiveness differences between consecutive scale intervals.

Given these requirements, the well-established ‘*Deck of Cards*’ method, based on the revised Simos’ procedure (Simos, 1990a, b), developed by Figueira and Roy (2002), and improved by Corrente et al. (2021; DCM-SRF), emerges as the most suitable alternative. This way, there is a finer discrimination between the different levels of the scale, which reliably and rigorously enhances the model’s discriminatory capacity and provides meaningfully quantified performance sub-indicators to be aggregated in the next step by means of a BoD approach.

On that account, two reference levels must first be defined to anchor the computations, such as 0 and 1 (Pereira & Pereira, 2023). Now adapting the steps of the DCM-SRF for the construction of interval scales provided by Pereira, Figueira, and Marques (2020), in line with Bottero et al. (2018), we must then:

1. Mind the discrete scale of a generic output  $y$ ,

$$S_y = l_1, \dots, l_i, \dots, l_d,$$

where  $l_1 < l_2 < \dots < l_i < \dots < l_d$  and ‘<’ denotes a ‘ranked strictly lower’ relation;

2. Set two reference levels  $l_k$  and  $l_q$  and attribute them utility values  $U(l_k) = 0$  and  $U(l_q) = 1$  due to their frequent use;
3. Query the DMs regarding the ranking of the levels and the placement of a number of blank cards  $c_i$  in the intervals between every two consecutive levels  $l_i$  and  $l_{i+1}$ , with  $i = 1, \dots, d - 1$ , such that:

$$l_1 c_1 \dots l_k c_k l_{k+1} c_{k+1} \dots l_i c_i l_{i+1} \dots l_{q-1} c_{q-1} l_q \dots l_{d-1} c_{d-1} l_d;$$



4. Compute the unit valuation

$$\alpha = \frac{U(l_q) - U(l_k)}{h},$$

where

$$h = \sum_{i=k}^{q-1} c_i + 1$$

by only taking into account the levels between  $l_k$  and  $l_q$ ;

5. Compute the utility value  $U(l_i)$  per level  $i$ , with  $i = 1, \dots, d$  as follows:

$$U(l_i) = \begin{cases} U(l_k) - \alpha \left( \sum_{f=i}^{k-1} c_f + 1 \right), & \text{for } i = 1, \dots, k - 1 \\ U(l_k) + \alpha \left( \sum_{f=k}^{i-1} c_f + 1 \right), & \text{for } i = k + 1, \dots, q - 1, \dots, q + 1, \dots, d \end{cases}$$

6. Iterate the procedure per indicator.

Since this procedure is suitable for discrete scales, as is the case of the ordinal scales of the Mobility Scoreboard, each of its categories corresponds to a level of 0% ('None of the elements exists'), 25% ('Systems fulfil only a limited part of the criteria analysed'), 50% ('Only some aspects are implemented'), 75% ('Most aspects appear in the system'), and 100% ('All criteria are fully met') within each indicator, given their maximisation intent, in the interest of connecting the aforementioned categories and levels  $l_1 < l_2 < \dots < l_i < \dots < l_d$ .

### 3.2.2 Step 2: The 'Benefit-of-the-Doubt' approach

#### Outline

To compute the CI pertaining to the 'Youth on the Move' learning mobility flagship initiative, we resort to a non-parametric technique entrenched in DEA. Although the definition of relative efficiency dates back to the concept introduced by Farrell (1957), it was not until the seminal work of Charnes et al. (1978) that its measurement was operationalised and put to the test. Still, its evolution towards the *BoD* approach is slightly more recent, culminating in the framework of Cherchye et al. (2007). In fact, the *BoD* formulation is equivalent to the formulation proposed by Charnes et al. (1978) assuming constant returns-to-scale (CRS), but with a unitary dummy input. It is precisely this input characteristic that allows it to manoeuvre each decision-making unit (DMU) in the direction of better performance.

After its mention by the Organization for Economic Cooperation and Development as a CI-building method (Nardo et al., 2008), *BoD*-based CIs have been finding applications in various sectors and computed using different variants of the standard *BoD* formulations to aggregate the individual sub-indicators. With respect to the empirical applications, they include corporate social responsibility (see, e.g. Oliveira et al. 2019), education (see, e.g. Camanho et al., 2023), energy (see, e.g. Zanella et al. 2015), environment (see, e.g. Zanella et al. 2013), health (see, e.g. Pereira et al. 2021), justice (see, e.g. Bogetoft & Wittrup 2021), mining (see, e.g. Oliveira et al. 2020), sustainability (see, e.g. Pereira & Marques 2022b), water (see, e.g. Sala-Garrido et al. 2021; Vilarinho et al. 2023), and waste (see, e.g. Rogge et al. 2017). As alternatives to the traditional *BoD* approach, several studies have proposed for example a multiplicative aggregation such as the geometric index approach (Van Puyenbroeck and

Rogge, 2017) and the generalized weighted averages (Rogge, 2018a), a procedure to aggregate individual composite indicators into group composite index (Rogge, 2018b) and models to account for non-compensability (see, e.g., Fusco, 2015, 2023).

If we consider the six Mobility Scoreboard key performance indicators in higher education as the vector of output indicators  $Y_j^t$ , the BoD approach computes the weights  $u_r$  linked to those outputs without *a priori* assumptions. This way, the weights are optimised to maximise the strengths of each European country—DMUs  $j$ , with  $j = 1, \dots, n$ ,—thus granting them the “benefit of the doubt” regarding each output per period  $t$  ( $t = 1, \dots, p$ ). The production possibility set (PPS)  $T^t$  of each period is enveloped by the frontier underpinning the comparison.

Following Van Puyenbroeck (2018), an input-oriented formulation of the BoD model is preferable due to its intuitiveness as a sole aggregator of outputs. However, despite the current setting being devoid of inputs, the incorporation of weight restrictions (justified in Section 3.2.2) requires the use of an output-oriented formulation, following Camanho et al. (2023). Model (1) mathematically expresses this determination for DMU 0 (the DMU under assessment):

$$[E_0^t(\mathbb{1}, Y_0^t)]^{-1} = \min \quad v \quad (1)$$

$$\text{subject to} \quad \sum_{r=1}^s u_r y_{r0}^t = 1 \quad (2)$$

$$\sum_{r=1}^s u_r y_{rj}^t - v \leq 0, \quad j = 1, \dots, n \quad (3)$$

$$u_r \geq 0, \quad r = 1, \dots, s \quad (4)$$

$$v \geq 0 \quad (5)$$

Weights  $u_r$  and  $v$  are the decision variables of Model (1), which aims to minimise the underperformance of the DMU under assessment (DMU 0) in period  $t$  under four constraints. First, the normality constraint given by Expression (2) dictates that the performance of DMU 0 is equal to 1, at best, when its optimal weights are considered. Hence,  $[E_0^t(\mathbb{1}, Y_0^t)]^{-1} = 1$  denotes a best-performing DMU and  $E_0^t(\mathbb{1}, Y_0^t) > 0$  denotes an underperforming DMU. Second, Expression (3) ensures that the performance of all DMUs in the sample is lower than or equal to one since there is a single unitary input. Third, the non-negativity constraints given by Expression (4) and (5) impose that weights  $u_r$  are greater or equal to 0, i.e., non-negative. Note that Model (1) intrinsically assumes CRS.

### Weight restrictions

If we look at Model (1) from the point of view of the education sector, we realise that there is an aspect being ignored in its formulation that cannot be overlooked—the preferences over the various learning mobility dimensions. This aspect finds an explanation in the distinct priorities of each nation, which are also a reflection of their own identity (D’Inverno & De Witte, 2020). Hence, weight restrictions must be included in Model (1).

Among the several types of weight restrictions that can be found in the literature (Cherchye, Moesen, Rogge, and van Puyenbroeck, 2007; Sarrico & Dyson, 2004), we adopt the favoured assurance region type I (ARI) ones in Expression (6) in line with Zanella et al. (2015) and Calabria et al. (2018). This way, it is possible to constrain the relative importance of each

Mobility Scoreboard indicator within a certain range, such that

$$\phi_r \leq \frac{u_r \bar{y}_{r0}}{\sum_{r=1}^s u_r \bar{y}_{r0}} \leq \psi_r, \quad r = 1, \dots, s, \quad (6)$$

with two weight restrictions imposed to each indicator  $r$  (lower and upper bound corresponding to  $\phi_r$  and  $\psi_r$ , respectively) and  $s + s$  constraints in total across all indicators. The restrictions resort to the use of an “artificial DMU” instead of being DMU-specific (Zanella, Camanho, and Dias, 2015).  $\bar{y}_r$  denotes the average value of indicator  $r$  observed for all DMUs in the sample (Camanho, Stumbriene, Barbosa, and Jakaitiene, 2023).

Bear in mind that we imposed weight restrictions on the lower and upper bounds of the relative importance attributed to each indicator to guarantee not only that all Mobility Scoreboard indicators are considered in the assessment, but also that a certain indicator would not assume overwhelming importance. These values were obtained via the DCM-SRF described above (Sect. 3.2.1), with the caveat that a 50% and a 150% normalisation were applied to its results to denote the lower and upper bounds, respectively. Indeed, rather than being used to convert an ordinal scale into an interval one, the DCM-SRF’s rationale can be used for computing “weights” as in the traditional MCDA sense. By resorting to the DecSpace platform,<sup>3</sup> it is possible to interactively generate those relative importance multipliers with the DM. Therefore, more realistic weight restrictions can be generated while adhering to the best practices in the literature (Camanho, Stumbriene, Barbosa, and Jakaitiene, 2023).

*Robust setting* The deterministic essence of the BoD approach suffers from estimation drawbacks in the presence of measurement errors or data outliers (De Witte & Schiltz, 2018). As a result, a robust version was created, inspired by the order- $m$  proposal of Cazals et al. (2002). This way, the influence of such obstacles is limited by means of re-sampling, i.e., by drawing with replacement  $m < n$  sub-samples from the original sample  $B$  times, less extreme benchmarks are considered and more *robust BoD* estimates are generated, following Expression (7):

$$\left[ \hat{E}_0^t(\mathbb{1}, Y_0^t) \right]^m = \frac{\sum_{b=1}^B \left[ \hat{E}_0^t(\mathbb{1}, Y_0^t) \right]^{b,m}}{B} \quad (7)$$

When  $m \rightarrow +\infty$ , the scores computed by the robust BoD model match the ones yielded by the deterministic model (Cazals, Florens, and Simar, 2002; Tauchmann, 2012). However, cases of super-performance may occur when the DMU under assessment is not a part of the re-sampled DMUs used to compute its own performance, inducing a DMU located above the frontier and, consequently, a performance score above 1. This means that a super-performing country can be regarded as a country with better performances than the average  $m$  other countries in the sub-sample. Naturally, the number of super-performing DMUs is related to the choice of  $m$  given the probability of DMU 0 not being a part of the sub-sample.

#### *Conditional setting*

Finally, when certain variables are not under the control of DMUs but do directly influence the production process, they need to be included in the calculation of robust BoD scores. Otherwise, some units may be unfairly assessed simply because their operating context is not taken into account. Despite the specificity of learning mobility and the lack of a clear sense of which exogenous factors are able to influence the performance of European countries in that regard, the literature has shown that data based on surveys significantly influences the attainable set (Verschelde and Rogge, 2012; Cordero, Salinas-Jiménez, and Salinas-Jiménez, 2017). As this is the case with the Mobility Scoreboard—given the qualitative information

<sup>3</sup> Available at <http://decspace.sysresearch.org/>.

collected through a survey (see Sect. 3.1) -, such *environmental variables*, denoted by vector  $Z_l^t = \{z_1^t, \dots, z_l^t, \dots, z_k^t\}$ , can be embedded in the model and generate conditional scores, transforming it into the *robust conditional BoD model* formulated in Expression (8):

$$\left[ \hat{E}_0^t(\mathbb{1}, Y_0^t) \right]^{m,z} = \frac{\sum_{b=1}^B \left[ \hat{E}_0^t(\mathbb{1}, Y_0^t) \right]^{b,m,z}}{B} \quad (8)$$

Be that as it may, there is no information available regarding respondent characteristics like in De Witte and Schiltz (2018). Hence, a measure of welfare, economic growth, and social development must be considered as a suitable and comprehensive environmental variable. The gross domestic product would be a possibility if not for its disregard for social development. Therefore, we saw fit to recognise the Human Development Index (HDI) as an appropriate alternative since it combines key dimensions of human development: health, education, and standard of living. The HDI was considered for 2016, 2019, and 2021.<sup>4</sup>

Now, unlike in Expression (7), where all DMUs have the same probability of being a part of the sub-sample, here the probability of a DMU being drawn depends on  $Z$ . To that end, a kernel function must be estimated to smooth this vector. Depending on whether  $Z$ 's variables are continuous or discrete, distinct approaches should be followed. See Bădin et al. (2010) for the former and Li and Racine (2004) for the latter. Yet, ultimately, the generated performance scores are not only robust to measurement errors and data outliers but also dependent on exogenous heterogeneity. Once again, cases of super-performing DMUs are a possibility (given scores larger than 1), but, this time, a super-performing country can be regarded as a country with better performances than the average  $m$  other countries in the sub-sample subject to similar contextual conditions.

At last, following the recommendation suggested by Henriques et al. (2022) when dealing with small samples, the value of  $m$  has been set equal to 36, i.e., the number of countries in each year.<sup>5</sup> This choice finds support in the work of Daraio and Simar (2007) since its value can be set by considering all possible DMUs in the sample.  $B$  was established as 2000 following the best practices in the field (see, e.g., De Witte & Schiltz 2018, Henriques et al., 2022).

### 3.2.3 Step 3: The global Malmquist productivity index

The relative efficiency measures computed by DEA-based techniques imply that the performances of the sampled DMUs are compared against each other. If the ultimate goal of efficiency measurement is to identify areas with potential for improvement at a given moment in time and for a sample of comparable DMUs, when considering trends in the evolution of efficiency over time, accounting for the possibility of occurring movements in the position of the frontier over time, we need to bring into play a method that is able to measure whether a DMU improved or declined from one period to the next one. This is especially relevant when a policy is being examined, which is the case of the 'Youth on the Move' flagship initiative.

As a result, we must introduce the Malmquist Productivity Index (*MPI*). Malmquist indices were introduced by Caves et al. (1982). The authors name these indices after Malmquist (1953), who had earlier proposed constructive input quantity indices as ratios of distance functions. Caves et al. (1982) defined an input-based productivity index relative

<sup>4</sup> Data for the 2022 HDI is not available yet.

<sup>5</sup> Note that the Euridyce Network divides Belgium into a French community, a German-speaking community, and a Flemish community, which, in itself, corresponds to a single country with 3 observations per reference year.

to a single technology corresponding to a given period  $t$  or  $t + 1$ . The  $MPI$  was treated as theoretical until its enhancement by Färe et al. (1994). A major contribution of these authors was to relax the efficiency assumption and provide DEA models for its calculation. In particular, Färe et al. (1994) defined an output-oriented productivity index as the geometric mean of the two Malmquist indices of Caves et al. (1982) referring to technology at period  $t$  and  $t + 1$ . However, the index of Färe et al. (1994) does not fulfil the circularity property, which is especially relevant when considering time series with more than two periods under assessment. Since we are analysing three periods, this property must hold. Therefore, we adopt the global  $MPI$  ( $GMPI$ ) proposed by Pastor and Lovell (2005). The  $GMPI$  combines the DMUs of all periods into a single global frontier, henceforth named *metafrontier*, which is considered as the reference to estimate the  $MPI$ . Although this concept was launched by Battese and Rao (2002), it was Battese et al. (2004) that set forth its deterministic non-parametric facet. Fundamentally, notwithstanding the fact that there is a PPS  $T^t$  per period  $t$ , a metatechnology  $T^M = \bigcup_{t=1}^P T^t$  envelops all PPSs. We point out that this set is conceptually similar to the corresponding sequential set at the final period, recently developed in a BoD framework by Walheer (2024), and the overall set by Afsharian and Ahn (2015).

However, despite the argument provided by Afsharian and Ahn (2015) that an overall Malmquist index should be based on a non-convex metatechnology to prevent non-homogeneity issues when determining the global benchmark technology, its computation toll proved to be troublesome. Hence, following Camanho et al. (2023), we adopt a convex metatechnology  $GMPI$  in line with Pastor and Lovell (2005). Lastly, we formulate the  $GMPI$  for DMU 0 as

$$GMPI_0^{t,t+1}(\mathbb{1}, Y_0^t, \mathbb{1}, Y_0^{t+1}) = \frac{\hat{E}_0^M(\mathbb{1}, Y_0^{t+1})}{\hat{E}_0^M(\mathbb{1}, Y_0^t)} \tag{9}$$

The interpretation of the  $GMPI$  is rather straightforward. If:

- $GMPI_0^{t,t+1} < 1$ , then the performance of DMU 0 declined from period  $t$  to period  $t + 1$ ;
- $GMPI_0^{t,t+1} = 1$ , then the performance of DMU 0 remained constant from period  $t$  to period  $t + 1$ ;
- $GMPI_0^{t,t+1} > 1$ , then the performance of DMU 0 improved from period  $t$  to period  $t + 1$ .

The adoption of a metafrontier appeals to a slender adaptation of the BoD formulation proposed in Model (1), as shown in Model (10):

$$\left[ E_0^M(\mathbb{1}, Y_0^t) \right]^{-1} = \min \quad v \tag{10}$$

$$\text{subject to} \quad \sum_{r=1}^s u_r y_{r0}^t = 1, \tag{11}$$

$$\sum_{r=1}^s u_r y_{rj}^t - v \leq 0, \quad \begin{cases} j = 1, \dots, n \\ t = 1, \dots, P \end{cases} \tag{12}$$

$$\frac{u_r \bar{y}_r}{\sum_{r=1}^s u_r \bar{y}_r} \geq \phi_r, \quad r = 1, \dots, s \tag{13}$$

$$\frac{u_r \bar{y}_r}{\sum_{r=1}^s u_r \bar{y}_r} \leq \psi_r, \quad r = 1, \dots, s \tag{14}$$

$$u_r \geq 0, \quad r = 1, \dots, s \tag{15}$$

$$v \geq 0, \quad (16)$$

It follows that the CI  $E_0^M(\mathbb{1}, Y_0^t)$  now corresponds to the inverse of the underperformance of DMU 0 in period  $t$  bearing in mind a metatechnology. Note that it is also generalisable to the performance of DMU 0 in period  $t + 1$ , coinciding with  $E_0^M(\mathbb{1}, Y_0^{t+1})$ .

From another angle, the *GMPI* is also typically decomposed into two components: *efficiency change (EC)* and *best-practice change (BPC)*. Expression (17) represents this decomposition.

$$\begin{aligned} GMPI_0^{t,t+1}(\mathbb{1}, Y_0^t, \mathbb{1}, Y_0^{t+1}) &= \frac{\hat{E}_0^{t+1}(\mathbb{1}, Y_0^{t+1})}{\hat{E}_0^t(\mathbb{1}, Y_0^t)} \cdot \left[ \frac{\hat{E}_0^M(\mathbb{1}, Y_0^{t+1}) \hat{E}_0^t(\mathbb{1}, Y_0^t)}{\hat{E}_0^{t+1}(\mathbb{1}, Y_0^{t+1}) \hat{E}_0^M(\mathbb{1}, Y_0^t)} \right] \\ &= EC_0^{t,t+1} \cdot BPC_0^{t,t+1} \end{aligned} \quad (17)$$

Essentially,  $EC_0^{t,t+1}$  denotes the change experienced by the distance of DMU 0 to the frontier between period  $t$  and period  $t + 1$ . If:

- $EC_0^{t,t+1} < 1$ , then DMU 0 is farther from the frontier in period  $t + 1$  than it was in period  $t$ ;
- $EC_0^{t,t+1} = 1$ , then DMU 0 remained at the same distance to the frontier in period  $t + 1$  as it was in period  $t$ ;
- $EC_0^{t,t+1} > 1$ , then DMU 0 is closer to the frontier in period  $t + 1$  than it was in period  $t$ .

$BPC_0^{t,t+1}$  denotes the change in the best practice gap in relation to the metafrontier between period  $t$  and period  $t + 1$  from the perspective of DMU 0. If:

- $BPC_0^{t,t+1} < 1$ , then the benchmark technology is farther from the global benchmark technology in period  $t + 1$  than it was in period  $t$ , according to the output mix of DMU 0 in periods  $t + 1$  and  $t$ , respectively;
- $BPC_0^{t,t+1} = 1$ , then the benchmark technology is at the same distance to the global benchmark technology in period  $t + 1$  as it was in period  $t$ , according to the output mix of DMU 0 in periods  $t + 1$  and  $t$ , respectively;
- $BPC_0^{t,t+1} > 1$ , then the benchmark technology is closer to the global benchmark technology in period  $t + 1$  than it was in period  $t$ , according to the output mix of DMU 0 in periods  $t + 1$  and  $t$ , respectively.

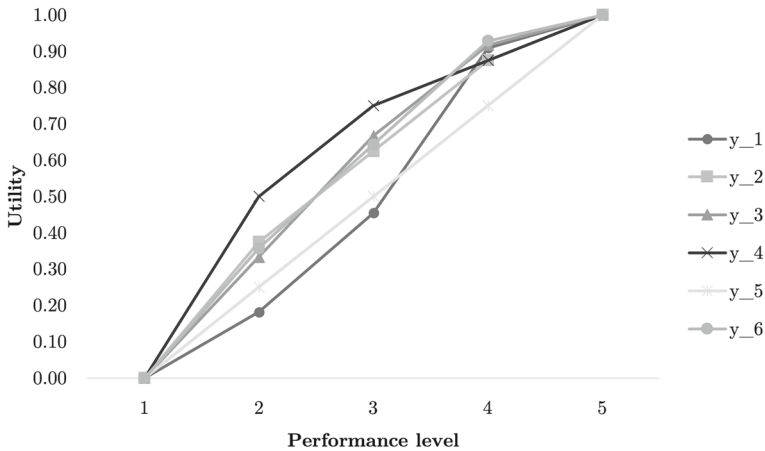
Bearing in mind the property of circularity mentioned above, the *GMPI* can be used to compare the three time periods being analysed here via

$$\begin{aligned} GMPI_0^{t,t+2}(\mathbb{1}, Y_0^t, \mathbb{1}, Y_0^{t+2}) &= GMPI_0^{t,t+1}(\mathbb{1}, Y_0^t, \mathbb{1}, Y_0^{t+1}) \cdot \\ &GMPI_0^{t+1,t+2}(\mathbb{1}, Y_0^{t+1}, \mathbb{1}, Y_0^{t+2}) \end{aligned} \quad (18)$$

in line with Camanho et al. (2023). This relationship also applies to the *EC* and the *BPC*.

## 4 Results and discussion

Here we start by addressing the incorporation of preferences in Sect. 4.1, before moving on to the assessment of performance in Sect. 4.2. Section 4.3 contains the analysis of performance change. Finally, an illustrative example of an underperforming country and its improvement potential in terms of education policy, based on our framework, is provided in Sect. 4.4.



**Fig. 1** Utility scales of the six Mobility Scoreboard thematic indicators

#### 4.1 Preference incorporation

The elicitation of the preferences of the DM—a well-known academia expert in European education economics—yielded the blank card structure displayed in Table 2, which, after using the DCM-SRF, returned the corresponding utility scale per indicator. Their comparison with the original values is shown in Table 3. Figure 1 graphically depicts these preferences. Note that  $y_2$ ,  $y_3$ ,  $y_4$ , and  $y_6$  present a concave shape denoting the DM’s risk-averse nature regarding ‘Foreign language preparation’, the ‘Portability of grants and loans’, the ‘Support to disadvantaged learners’, and the ‘Recognition of qualifications’, respectively. This means that the utility scales of these four indicators denote greater differences between lower performance levels than between higher performance levels. Interestingly, on the one hand,  $y_5$  resulted in a linear preference structure since the DM did not consider it adequate to attribute greater or lower importance to any particular performance level when recognising learning outcomes. On the contrary,  $y_1$  became a hybrid scale—convex at first and concave at last—given the DM’s lower preference for lower performance levels of ‘Information and guidance’, but higher preference for greater performance levels of that indicator.

Regarding the weight restrictions, as described in Section 3.2.2, we resorted to the same MCDA method, but, this time, to compute weights as if we were in a traditional MCDA process. By applying a 50% and a 150% normalisation to those weights, we obtained the results shown in Table 4.

#### 4.2 Performance assessment

The CI aggregating the six thematic indicators of the Mobility Scoreboard in the areas of information and guidance, foreign language preparation, portability of grants and loans, support for disadvantaged learners, recognition of learning outcomes through the ECTS, and recognition of qualifications in the EHEA for the 36 sampled European countries and macro-regions in the three reference academic years (2015/2016, 2018/2019, and 2022/2023) taking into account their human development context is shown in Table 5.



**Table 2** Utility conversion procedure

Level	Performance		Indicator		y <sup>1</sup>		y <sup>2</sup>		y <sup>3</sup>		y <sup>4</sup>		y <sup>5</sup>		y <sup>6</sup>	
	Blank card(s)	Interval scale	Blank card(s)	Interval scale	Blank card(s)	Interval scale	Blank card(s)	Interval scale	Blank card(s)	Interval scale	Blank card(s)	Interval scale	Blank card(s)	Interval scale	Blank card(s)	Interval scale
None of the elements exist	1	0.00	2	0.00	3	0.00	3	0.00	3	0.00	0	0.00	4	0.00	4	0.00
Systems fulfil only a limited part of the criteria analysed	2	0.18	1	0.38	3	0.33	1	0.50	0	0.25	0	0.36	3	0.25	3	0.36
Only some aspects are implemented	4	0.45	1	0.63	2	0.67	0	0.75	0	0.50	0	0.64	3	0.50	3	0.64
Most aspects appear in the system	0	0.91	0	0.88	0	0.92	0	0.88	0	0.75	0	0.93	0	0.75	0	0.93
All criteria are fully met	5	1.00	–	1.00	–	1.00	–	1.00	–	1.00	–	1.00	–	1.00	–	1.00

**Table 3** Descriptive statistics of the original and converted performance values

Reference year	Indicator	Original		Converted				
		Median	Minimum	Maximum	Mean	Standard deviation	Minimum	Maximum
2015/2016	y1	3	1	5	0.46	0.39	0.00	1.00
	y2	4	1	5	0.74	0.25	0.00	1.00
	y3	3.5	1	5	0.67	0.37	0.00	1.00
	y4	3	1	5	0.63	0.32	0.00	1.00
	y5	4	1	5	0.56	0.38	0.00	1.00
	y6	3	1	5	0.63	0.31	0.00	1.00
2018/2019	y1	3	1	5	0.52	0.40	0.00	1.00
	y2	4	1	5	0.73	0.25	0.00	1.00
	y3	4	1	5	0.69	0.36	0.00	1.00
	y4	2	1	5	0.53	0.27	0.00	1.00
	y5	3	2	5	0.62	0.24	0.25	1.00
	y6	3	1	5	0.67	0.30	0.00	1.00
2022/2023	y1	3	1	5	0.56	0.37	0.00	1.00
	y2	4	1	5	0.74	0.26	0.00	1.00
	y3	4	1	5	0.72	0.34	0.00	1.00
	y4	2	1	5	0.53	0.26	0.00	1.00
	y5	4	2	5	0.65	0.24	0.25	1.00
	y6	4	2	5	0.76	0.26	0.36	1.00

**Table 4** Values of the weight restrictions

Indicator	Original DCM-SRF weight	50% normalised DCM-SRF weight ( $\phi_r$ )	150% normalised DCM-SRF weight ( $\psi_r$ )
$y_1$	16.00%	8.00%	24.00%
$y_2$	11.11%	5.56%	16.67%
$y_3$	19.78%	9.89%	29.67%
$y_4$	19.78%	9.89%	29.67%
$y_5$	11.11%	5.56%	16.67%
$y_6$	22.22%	11.11%	33.33%
<b>Total</b>	<b>100.00%</b>	<b>50.00%</b>	<b>150.00%</b>

Despite the great variability of results among the three periods, with several countries improving and others deteriorating their performances, it is clear that the standard deviation constantly decreased from 2015/2016 to 2022/2023 while the average ended up increasing despite a lower value in 2018/2019. In particular, although no country retained their status as a top performer across the three periods, only Belgium's Flemish community exhibited an average greater than one between 2015/2016 and 2022/2023 (1.0016), followed closely by Finland (0.9954), Romania (0.9915), and Hungary (0.9905). On the contrary, Bulgaria was, on average, the worst performer between 2015/2016 and 2022/2023 (0.5150), despite a late improvement in 2022/2023, with Serbia revealing a similar behaviour (0.5284). Overall, in the absence of a single composite measure, if we look at the bigger picture of the disaggregate results of the European Education and Culture Executive Agency (2023), ours are aligned with theirs. For instance, Belgium's Flemish community, Finland, and France remain as the top performers and Bulgaria and Serbia remain as the bottom performers. Nonetheless, there are some differences in the sense that countries that perform relatively well overall according to the European Education and Culture Executive Agency (2023) - e.g., Luxembourg, Malta, the Netherlands, Austria, and Norway -, as well as countries that need further policy development - e.g., Bosnia and Herzegovina - revealed a different reality than what was reported, i.e., on average, the former perform in Q2/Q3 rather than in Q1 and the latter in Q2 rather than in Q4 (assuming we divide the performance scores into quartiles). We recall that the discussed results have been obtained by applying a 50%-normalisation. The resulting CI scores are not sensitive to the lower and upper weight bounds; for instance, 25%, 40%, and 45% normalisations have been additionally tested.

To test whether there are geographical and/or economic differences, we resorted to non-parametric hypothesis tests (see, e.g., Pereira et al. 2023). Therefore, the Kruskal-Wallis  $H$  test was applied to the sampled countries, grouped according to *Regions of Europe*,<sup>6</sup> and the Mann-Whitney  $U$  test was applied to the sampled countries, grouped according to the *Type of economic relation with the EU*,<sup>7</sup> to verify the existence of statistically significant differences between their performance scores. For a significance level of 5%, the first test retained the null hypothesis ( $p$ -value = 0.566), as did the second test ( $p$ -value = 0.639). Thus, the *Regions of Europe* and the *Type of economic relation with the EU* do not influence the learning mobility performance scores of European countries.

<sup>6</sup> The considered European regions are either Southern Europe, Central Europe, Northern Europe, or Eastern Europe.

<sup>7</sup> The considered type of economic relation of countries with the EU is either being an EU member or being a member of the EEA/EFTA or an EU candidate.

Table 5 CI scores of the Mobility Scoreboard

Country	2015/2016		2018/2019		2022/2023	
	Performance score	Ranking	Performance score	Ranking	Performance score	Ranking
Belgium - French community	0.6725	31	0.5572	33	0.7416	30
Belgium - German-speaking community	0.9743	7	0.9075	14	0.9075	17
Belgium - Flemish community	1.0059	1	1.0078	1	0.9910	5
Bulgaria	0.3915	36	0.5099	35	0.6435	32
Czechia	0.8332	27	0.7739	24	0.8401	22
Denmark	0.9330	11	0.9257	10	0.9268	15
Germany	0.9914	5	0.9568	7	0.9536	10
Estonia	0.8432	26	0.8288	22	0.8773	20
Ireland	0.6473	32	0.5675	32	0.5693	36
Greece	0.7258	28	0.6472	29	0.8401	23
Spain	0.5579	33	0.5494	34	0.5862	35
France	0.9529	8	1.0076	2	0.9903	6
Croatia	0.9261	12	0.9170	11	1.0026	1
Italy	0.8739	20	0.8811	18	0.8843	19
Cyprus	0.9443	9	0.6650	28	0.7885	27
Latvia	0.4204	35	0.6190	31	0.7497	28
Lithuania	0.8659	21	0.9118	13	0.9723	9
Luxembourg	0.8648	22	0.8299	21	0.8310	24
Hungary	1.0018	2	0.9919	5	0.9780	7
Malta	0.8830	19	0.9001	15	0.9189	16
Netherlands	0.9422	10	0.9354	8	0.9348	13
Austria	0.8522	24	0.8762	19	0.9746	8

Table 5 continued

Country	2015/2016		2018/2019		2022/2023	
	Performance score	Ranking	Performance score	Ranking	Performance score	Ranking
Poland	0.9964	3	0.8847	17	0.9455	12
Portugal	0.9071	14	0.6309	30	0.7353	31
Romania	0.9929	4	0.9838	6	0.9979	3
Slovenia	0.7232	29	0.7391	25	0.6063	34
Slovakia	0.8968	17	0.7740	23	0.8292	25
Finland	0.9849	6	1.0070	3	0.9943	4
Sweden	0.8433	25	0.8927	16	0.8929	18
Bosnia and Herzegovina	0.9009	16	0.9119	12	0.9482	11
Iceland	0.7196	30	0.7312	27	0.8066	26
Liechtenstein	0.9163	13	0.8725	20	0.8724	21
Montenegro	0.9047	15	0.7333	26	0.7439	29
Norway	0.8870	18	0.9301	9	0.9293	14
Serbia	0.4808	34	0.4723	36	0.6320	33
Türkiye	0.8589	23	0.9953	4	0.9985	2
<b>Average</b>	<b>0.8366</b>	-	<b>0.8146</b>	-	<b>0.8565</b>	-
<b>Standard deviation</b>	<b>0.1633</b>	-	<b>0.1591</b>	-	<b>0.1293</b>	-
<b>Minimum</b>	<b>0.3915</b>	-	<b>0.4723</b>	-	<b>0.5693</b>	-
<b>Maximum</b>	<b>1.0059</b>	-	<b>1.0078</b>	-	<b>1.0026</b>	-

**Table 6** Key descriptive statistics of the relative importance of each Mobility Scoreboard indicator

Statistic	Indicator					
	y1	y2	y3	y4	y5	y6
Average	13.61%	13.17%	21.19%	17.63%	10.90%	23.50%
Standard deviation	6.71%	4.47%	8.46%	8.32%	4.95%	9.23%
Minimum	8.00%	5.56%	9.89%	9.89%	5.56%	11.11%
Maximum	24.00%	16.67%	29.67%	29.67%	16.67%	33.33%

**Table 7** Difference between the average BoD “weights” and the original DCM-SRF weights

Statistic	Indicator					
	y1	y2	y3	y4	y5	y6
Average BoD “weight”	13.61%	13.17%	21.19%	17.63%	10.90%	23.50%
Original DCM-SRF weight	16.00%	11.11%	19.78%	19.78%	11.11%	22.22%
Difference	-2.39%	+2.06%	+1.41%	-2.15%	-0.21%	+1.28%

If we focus on the peers, on the one hand, we find that European countries resort to at least 3 (Belgium’s French community in 2015/2016 and Austria in 2022/2023) and up to 20 (Slovakia in 2018/2019) peers. 8 is the average number of peers. On the other hand, some countries are never used as peers (Belgium’s French community, Ireland, Spain, and Iceland), whereas others are used by almost all countries (France is a peer 98 times). 24 is the average number of times a country acts as a peer. For the sake of space, we omit them from the paper.

From another angle, it is noteworthy to take a look at the relative importance per indicator in the construction of the CI (see Table 6). The optimised multipliers returned by the weight-restricted robust conditional BoD model indicate that the ‘Recognition of qualifications’ is the indicator with the highest average relative importance (23.50%), with the ‘Portability of grants and loans’ (21.19%) emerging as a close second place - both above their original DCM-SRF weight (22.22% and 19.78%, respectively) -, unlike the ‘Support to disadvantaged learners’ (17.63%, despite having the same original DCM-SRF weight as the latter indicator). The ‘Recognition of learning outcomes’ is the least important indicator (10.90%). Interestingly, the indicators with the highest relative importance denoted the highest standard deviation and *vice versa*. Additionally, it is also alluring to consider the differences between the average “weights” and the original MCDA weights (see Table 7). While some indicators are relatively less important after being optimised ( $y_1$ ,  $y_4$ , and  $y_5$ ), others are relatively more important ( $y_2$ ,  $y_3$ , and  $y_6$ ).

When we compare these results with the ones reported by the European Education and Culture Executive Agency (2023), we find a few differences, though. Our model returned a higher relative importance of the ‘Recognition of qualifications’, the ‘Portability of grants and loans’, and the ‘Support to disadvantaged learners’ on the computation of the CI. The ‘Recognition of learning outcomes’ and ‘Foreign language preparation’ were the indicators highlighted by the European Education and Culture Executive Agency (2023) as the ones in which European education systems fared better. Nevertheless, the report also painted a rather positive picture regarding the ‘Recognition of qualifications’ and the ‘Portability of grants and loans’.

### 4.3 Performance change

Table 8 contains the results of the *GMPI* and its components (the *EC* and the *BPC*) between the first and last period of analysis as well as between intermediate periods, computed following Expression (9) and (17). On average, European countries experienced a slight performance improvement (1.0503) from 2015/2016 to 2022/2023. This was accompanied not only by an average approximation of countries to the frontier between the two considered school years (1.0538), but also by an average distancing of benchmark technologies away from the global benchmark technology (0.9977). If not for the disruptive effect of the COVID-19 pandemic on mobility, perhaps these changes would have been different. Furthermore, in particular:

- Approximately 42% of the sampled nations experienced a performance decline in this period, with Portugal (0.8106) emerging as the one whose performance deteriorated the most - just Montenegro and Slovenia experienced a similar decline (0.8223 and 0.8384, respectively). Inversely, Latvia exhibited the highest performance growth (1.7834), with only Serbia resembling such an improvement (1.3144).
- About 47% of the European countries moved farther from the frontier in this period, being Montenegro the farthest from the frontier in 2022/2023 (0.8073). On the contrary, Latvia also manifested the highest *EC* (1.8787). No country comes near each of these countries' efficiency change scores.
- At last, around 53% of the benchmark technologies of the 36 EU Member States, EEA/EFTA countries, and EU candidates and macro-regions moved away from the global benchmark technology between 2015/2016 and 2022/2023. Bosnia and Herzegovina displayed the lowest *BPC* (0.8152) and Bulgaria the highest (1.1859). Once again, no country approaches these countries' best practice change levels.

If we compare these results with the ones reported by the European Education and Culture Executive Agency (2023), the alignment is even greater than in terms of performance assessment. In fact, Latvia and Austria have shown more significant progress than the majority of countries. This is due to their newly adopted policy measures regarding learning mobility in the education sector. Furthermore, Belgium's French community, Estonia, Greece, Croatia, Lithuania, Malta, Norway, and Serbia have all experienced a global *MPI* > 1, denoting their performance improvement to a Q1/Q2 status.

Moreover, we can resort to the Kruskal-Wallis *H* and Mann-Whitney *U* tests once more to verify the existence of statistically significant differences between the performance change scores of the grouped European countries. For the same significance level, both tests retained the null hypothesis ( $p$ -value = 0.228 and  $p$ -value = 0.325, respectively). Thus, the *Regions of Europe* and the *Type of economic relation with the EU* do not influence the learning mobility performance change scores of European countries.

Since important changes may have occurred between 2018/2019 and 2022/2023 due to the COVID-19 pandemic, we explore the evolution of performance between the first two periods and then between the second and the last period in the second part of Table 8. The overall average productivity gain (1.0503) observed between 2015/2016 and 2022/2023 is mostly driven by the change between 2018/2019 and 2022/2023 (1.0660), and in particular by the *EC* component (1.0818) as opposed to a decline in the *BPC* component (0.9857). On the one hand, this suggests that overall countries have reduced the distance from the best practice frontier in recent years, meaning that the learning mobility performance of the countries is becoming more homogeneous (Camanho, Stumbriene, Barbosa, and Jakaitiene, 2023). On the other hand, their distance to the benchmark technology has on average increased. This result can



**Table 8** Evolution of countries' performance between 2015/2016 and 2022/2023

Country	2015/2016–2022/2023			2015/2016–2018/2019			2018/2019–2022/2023			
	<i>GMPI</i>	<i>EC</i>	<i>BPC</i>	Ranking	<i>GMPI</i>	<i>EC</i>	<i>BPC</i>	<i>GMPI</i>	<i>EC</i>	<i>BPC</i>
Belgium - French Community	1.1028	1.1304	0.9756	9	0.8285	0.8058	1.0283	1.3309	1.4028	0.9488
Belgium German-speaking Community	0.9315	0.9357	0.9955	30	0.9314	0.9157	1.0172	1.0001	1.0218	0.9787
Belgium - Flemish Community	0.9852	0.9950	0.9901	24	1.0018	0.9965	1.0054	0.9834	0.9986	0.9848
Bulgaria	1.6437	1.3861	1.1859	2	1.3025	1.1883	1.0961	1.2620	1.1664	1.0819
Czechia	1.0083	0.9526	1.0585	20	0.9288	0.8790	1.0566	1.0857	1.0837	1.0018
Denmark	0.9934	1.0015	0.9919	22	0.9922	0.9743	1.0184	1.0012	1.0279	0.9740
Germany	0.9619	0.9723	0.9893	26	0.9651	0.9591	1.0063	0.9967	1.0138	0.9831
Estonia	1.0405	1.0278	1.0123	16	0.9830	0.9818	1.0012	1.0585	1.0469	1.0111
Ireland	0.8795	0.9047	0.9721	32	0.8768	0.8712	1.0064	1.0031	1.0385	0.9659
Greece	1.1575	1.1489	1.0075	5	0.8917	0.8658	1.0299	1.2981	1.3269	0.9783
Spain	1.0507	1.0302	1.0199	13	0.9847	0.9302	1.0586	1.0670	1.1076	0.9634
France	1.0392	1.0050	1.0341	17	1.0574	1.0107	1.0462	0.9828	0.9943	0.9884
Croatia	1.0827	1.1419	0.9481	10	0.9902	0.9986	0.9916	1.0934	1.1435	0.9562
Italy	1.0119	0.9721	1.0410	18	1.0083	0.9512	1.0600	1.0036	1.0219	0.9820
Cyprus	0.8351	0.8870	0.9414	34	0.7043	0.7258	0.9703	1.1857	1.2221	0.9702
Latvia	1.7834	1.8787	0.9493	1	1.4724	1.4906	0.9878	1.2112	1.2603	0.9610
Lithuania	1.1228	1.1103	1.0113	7	1.0530	1.0659	0.9878	1.0664	1.0416	1.0237
Luxembourg	0.9609	0.9547	1.0065	27	0.9596	0.9374	1.0238	1.0013	1.0185	0.9831
Hungary	0.9763	0.9995	0.9768	25	0.9901	1.0028	0.9873	0.9860	0.9967	0.9893
Malta	1.0407	0.9795	1.0625	15	1.0194	0.9489	1.0743	1.0209	1.0323	0.9890
Netherlands	0.9921	0.9972	0.9949	23	0.9927	0.9810	1.0119	0.9994	1.0165	0.9832
Austria	1.1436	1.1614	0.9846	6	1.0282	1.0159	1.0121	1.1122	1.1433	0.9728
Poland	0.9490	0.8853	1.0719	29	0.8879	0.8436	1.0526	1.0688	1.0495	1.0184
Portugal	0.8106	0.8565	0.9465	36	0.6955	0.7133	0.9751	1.1656	1.2008	0.9707
Romania	1.0050	0.9879	1.0173	21	0.9908	0.9764	1.0148	1.0143	1.0118	1.0025
Slovenia	0.8384	0.8490	0.9875	33	1.0221	0.9914	1.0309	0.8203	0.8564	0.9579

Table 8 continued

Country	2015/2016–2022/2023			Ranking	2015/2016–2018/2019			2018/2019–2022/2023		
	<i>GMPI</i>	<i>EC</i>	<i>BPC</i>		<i>GMPI</i>	<i>EC</i>	<i>BPC</i>	<i>GMPI</i>	<i>EC</i>	<i>BPC</i>
Slovakia	0.9246	1.0350	0.8933	31	0.8631	0.9318	0.9263	1.0712	1.1108	0.9644
Finland	1.0095	1.0035	1.0060	19	1.0224	0.9980	1.0244	0.9874	1.0055	0.9820
Sweden	1.0588	1.0695	0.9900	11	1.0587	1.0519	1.0064	1.0002	1.0167	0.9837
Bosnia and Herzegovina	1.0526	1.2912	0.8152	12	1.0122	1.2019	0.8422	1.0399	1.0743	0.9680
Iceland	1.1210	1.0986	1.0204	8	1.0161	0.9824	1.0343	1.1032	1.1183	0.9865
Liechtenstein	0.9520	0.9352	1.0179	28	0.9521	0.9221	1.0326	0.9998	1.0143	0.9858
Montenegro	0.8223	0.8073	1.0186	35	0.8106	0.8189	0.9899	1.0144	0.9858	1.0290
Norway	1.0477	1.0484	0.9993	14	1.0485	1.0323	1.0157	0.9992	1.0156	0.9839
Serbia	1.3144	1.2677	1.0368	3	0.9822	0.9488	1.0351	1.3382	1.3361	1.0016
Türkiye	1.1626	1.2279	0.9468	4	1.1588	1.2015	0.9645	1.0032	1.0220	0.9816
Average	1.0503	1.0538	0.9977		0.9856	0.9753	1.0117	1.0660	1.0818	0.9857
Standard deviation	0.1946	0.1907	0.0575		0.1380	0.1398	0.0439	0.1106	0.1136	0.0247
Minimum	0.8106	0.8073	0.8152		0.6955	0.7133	0.8422	0.8203	0.8564	0.9488
Maximum	1.7834	1.8787	1.1859		1.4724	1.4906	1.0961	1.3382	1.4028	1.0819

be attributed to the pandemic's disruption not only of teaching, learning, and research within individual countries but also various international activities among universities, particularly the physical mobility of students and staff (Li & Ai, 2022). However, this disruption has created opportunities for innovation and led to new learning mobility patterns such as virtual exchanges, collaborative online learning, and blended mobility (European Commission and Directorate-General for Education, Youth, Sport and Culture, 2021). The concept of "virtual mobility" has been valued and defined as "a form of academic mobility in which students and teachers in higher education can study or teach by using digital tools and platforms without physically travelling to another higher education institution abroad."<sup>8</sup> The global challenges of the pandemic, the rapid digitalisation, and the development of flexible learning mobility formats have led the European Commission to update the existing European learning mobility framework: from the 'Youth on the Move' to the 'Europe on the Move', to address learners from all groups, teachers, and education staff as well as to promote more sustainable mobility, published on November 15, 2023 as a proposal for a Council Recommendation.<sup>9</sup> Also the existing Mobility Scoreboard is expected to be adjusted. The Council wants to build on existing initiatives to enhance the evidence base on learning mobility by "revamping the Mobility Scoreboard, in close cooperation with experts from the Member States, to follow up the implementation of this Recommendation and expand it to cover all education and training, and youth sectors" (ibidem). To this extent, the tool proposed in this paper is flexible enough to monitor the change and measure the impact of this new Recommendation when the Mobility Scoreboard is updated.

#### 4.4 Target practice

Finally, for the sake of performance improvement to reach the best practice frontier, underperforming countries must have targets to lead them towards their objective. This can be achieved by means of the dual - envelopment - formulation of Model (10). Its use is put into practice via the DMU that deteriorated its performance the most between 2015/2016 and 2022/2023 - Portugal. This country is used as an illustrative case for the policy implications that can be drawn from the quantitative performance assessment.

First, Portugal experienced the most significant performance decline in the sample from 2015/2016 to 2022/2023, remaining in Q4 in the last two reference years. In particular, on the one hand, the country exhibited the highest average distance increase to the frontier from the first to the last period; on the other hand, only Bosnia and Herzegovina, Slovakia, and Cyprus's benchmark technology were farther away from the global benchmark technology from the first to the last period. Second, regarding its peers, Portugal should (among other less predominant nations) look up to Poland, France, and Malta in 2015/2016, France, Italy, and Lithuania in 2018/2019, and France, Italy, and Czechia in 2022/2023. Although these nations were not top performers, on average, they returned performance scores in Q1 and/or Q2. Third, the relative importance of Portugal's indicators remained relatively stable across all periods, despite a progressive decrease in terms of the 'Recognition of learning outcomes' - the indicator with the lowest relative importance. In fact, the country's low performances seem to be explained by low-performance values in indicators with high relative importance and *vice versa*; for instance, the average relative importance Portugal places on 'Portability of grants and loans' is 68.8% lower than the average BoD "weight" shown in Table 6. In the

<sup>8</sup> Available at <https://epale.ec.europa.eu/en/content/invitation-moocs-open-virtual-mobility-learning-hub>.

<sup>9</sup> Available at <https://erasmus-plus.ec.europa.eu/document/proposal-for-a-council-recommendation-europe-on-the-move-learning-mobility-for-everyone>.

end, the country's education policy should look at France's example and aim at improvements in the area of 'Portability of grants and loans' by assessing and streamlining bureaucratic processes (e.g., via collaborative initiatives with other European countries), promoting a more open and accessible mobility environment for students (possible attached to an increase in funding), and implementing comprehensive information campaigns to raise awareness among students about the available support for international study.

## 5 Conclusion

We proposed a weight-restricted robust conditional BoD model hybridised with MCDA's DCM-SRF to incorporate the preferences of a DM specialised in education policy and economics into a framework capable of assessing the performance and evaluating the performance change of 36 European countries and macro-regions regarding their learning mobility education policies between 2015/2016 and 2022/2023. This unprecedented empirical application on the subject makes several contributions to the operations research literature (e.g., the use of the DCM-SRF to convert data and compute weight restrictions to be used in the BoD model), as well as the education literature (e.g., the absence of studies about learning mobility).

In general, from 2015/2016 to 2022/2023, European countries experienced modest performance growth, which may be indicative of the relatively static nature of the learning mobility policy area over the past eight years along with the disruptive effects of the pandemic. Indeed, more than 80% of these countries found themselves in a similar situation in 2022/2023 as they were in 2015/2016, within  $\pm 1$  standard deviation from the average *GMPI*. Notably, belonging to a specific European region or having a certain type of economic relation with the EU did not significantly influence a country's learning mobility performance or performance change, suggesting that localised policy interventions tailored to specific regional challenges could enhance performance.

Areas requiring further enhancement include the 'Recognition of learning outcomes', 'Foreign language preparation', and 'Information and guidance'. The former two areas might benefit from revising current measures to provide a more detailed examination of each country's progress towards full compliance. In contrast, 'Information and guidance' may not fully reflect recent incremental changes nor the systematic lack of attention to its provision. More in general, the proposed tool is flexible to adjust to the updated learning mobility framework and the revamped Mobility Scoreboard foreseen by the new Recommendation proposal (November 2023) drafted as a reply to the challenges brought by the pandemic, the rapid digitalisation, online learning and virtual mobility.

Policy recommendations should focus on better aligning with the European education policy framework, adhering to the Recommendation of the Council of the European Union (2011) for promoting automatic mutual recognition and qualification arrangements, and enhancing transparent communication among education stakeholders. Despite these efforts, the European Education and Culture Executive Agency (2023) notes the ongoing lack of top-level monitoring concerning 'Information and guidance', 'Support to disadvantaged learners', and 'Recognition of qualifications'.

Our analysis, constrained by the qualitative nature of data from the Eurydice Network, could have been enriched with additional information to integrate evidence more effectively from the questionnaire responses. Furthermore, the availability of regional-level data, especially in countries with decentralised education policies, would better capture local nuances in

policy implementation and its impacts on educational outcomes. Such enhancements would likely align our findings more closely with those reported by the European Education and Culture Executive Agency (2023).

Looking ahead, future research should adapt this framework to national and regional contexts at both the country and university levels, potentially offering transformative insights into the quality of learning mobility. Another promising direction could involve applying the method of Aparicio and Santin (2018) to assess group performance over time, which might provide deeper insights into the efficiency change and the best-practice change. However, integrating this method within an *MPI* framework presents challenges that need to be addressed to enhance the robustness and applicability of performance assessments in various contexts. Additionally, exploring the integration of artificial intelligence and machine learning techniques to predict future performance in learning mobility policies presents a promising avenue, reflecting the increasing influence of these technologies in science and society.

**Acknowledgements** We would like to thank the Foundation for Science and Technology (FCT, from the Portuguese abbreviation of *Fundação para a Ciência e a Tecnologia*) for supporting this research via the project ‘eduBEST: Education Systems Benchmarking with Frontier Techniques’ (2022.08686.PTDC). Miguel Alves Pereira would also like to thank the FCT for supporting this research via project UIDB/00097/2020. Furthermore, we would like to acknowledge Tommaso Agasisti for his invaluable contributions to this paper. We would also like to leave a word of appreciation to the two anonymous referees for their inestimable insights that improved earlier versions of this paper. The authors also appreciate the comments of John Walden, Paul Rouse, Mette Asmild, and Valentin Zelenyuk in previous drafts of this work. Our views (and any errors) are our own responsibility.

**Author Contributions** Not applicable.

**Funding** Open access funding provided by FCTIFCCN (b-on). This work was supported by the FCT (Grants No. 2022.08686.PTDC and UIDB/00097/2020).

**Availability of data and materials** Not applicable.

**Code availability** Not applicable.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Afsharian, M., & Ahn, H. (2015). The overall Malmquist index: A new approach for measuring productivity changes over time. *Annals of Operations Research*, 226(1), 1–27. <https://doi.org/10.1007/s10479-014-1668-5>
- Agasisti, T. (2011). Performances and spending efficiency in higher education: A European comparison through non-parametric approaches. *Education Economics*, 19(2), 199–224. <https://doi.org/10.1080/09645290903094174>
- Agasisti, T. (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education*, 49(4), 543–557. <https://doi.org/10.1111/ejed.12069>
- Agasisti, T., Munda, G., & Hippe, R. (2019). Measuring the efficiency of European education systems by combining data envelopment analysis and multiple-criteria evaluation. *Journal of Productivity Analysis*, 51(2–3), 105–124. <https://doi.org/10.1007/s11123-019-00549-6>
- Ahec Sonje, A., Deskar-Skrbic, M., & Sonje, V. (2018). Efficiency of public expenditure on education: Comparing Croatia with other NMS. *12th International Technology, Education and Development Conference* (pp. 2317–2326). Valencia, Spain. <http://library.iated.org/view/AHECSONJE2018EFF>
- Aparicio, J., Ortiz, L., & Santín, D. (2021). Comparing group performance over time through the Luenberger productivity indicator: An application to school ownership in European countries. *European Journal of Operational Research*, 294(2), 651–672.
- Aparicio, J., Perelman, S., & Santín, D. (2022). Comparing the evolution of productivity and performance gaps in education systems through deal: An application to Latin American countries. *Operational Research*, pp. 1–35.
- Aparicio, J., & Santín, D. (2018). A note on measuring group performance over time with pseudo-panels. *European Journal of Operational Research*, 267, 227–235. <https://doi.org/10.1016/j.ejor.2017.11.049>
- Arbona, A., Giménez, V., López-Estrada, S., & Prior, D. (2022). Efficiency and quality in Colombian education: An application of the metafrontier malmquist-luenberger productivity index. *Socio-Economic Planning Sciences*, 79, 101122.
- Bas, M. C., & Carot, J. M. (2022). A model for developing an academic activity index for higher education instructors based on composite indicators. *Educational Policy*, 36(5), 1108–1134. <https://doi.org/10.1177/0895904820951123>
- Battese, G. E., & Rao, D. S. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*, 1(2), 87–93.
- Battese, G. E., Rao, D. S. P., & O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, 21(1), 91–103. <https://doi.org/10.1023/B:PROD.0000012454.06094.29>
- Bogetoft, P., Heinesen, E., & Tranæs, T. (2015). The efficiency of educational production: A comparison of the Nordic countries with other OECD countries. *Economic Modelling*, 50, 310–321. <https://doi.org/10.1016/j.econmod.2015.06.025>
- Bogetoft, P., & Witttrup, J. (2021). Benefit-of-the-doubt approach to workload indicators: Simplifying the use of case weights in court evaluations. *Omega*, 103, 102375. <https://doi.org/10.1016/j.omega.2020.102375>
- Bottero, M., Ferretti, V., Figueira, J. R., Greco, S., & Roy, B. (2018). On the Choquet multiple criteria preference aggregation model: Theoretical and practical insights from a real-world application. *European Journal of Operational Research*, 271(1), 120–140. <https://doi.org/10.1016/j.ejor.2018.04.022>
- Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research*, 201(2), 633–640. <https://doi.org/10.1016/j.ejor.2009.03.038>
- Calabria, F. A., Camanho, A. S., & Zanella, A. (2018). The use of composite indicators to evaluate the performance of Brazilian hydropower plants. *International Transactions in Operational Research*, 25(4), 1323–1343. <https://doi.org/10.1111/itor.12277>
- Camanho, A. S., Stumbriene, D., Barbosa, F., & Jakaitiene, A. (2023). The assessment of performance trends and convergence in education and training systems of European countries. *European Journal of Operational Research*, 305(1), 356–372. <https://doi.org/10.1016/j.ejor.2022.05.048>
- Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6), 1393. <https://doi.org/10.2307/1913388>
- Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106(1), 1–25. [https://doi.org/10.1016/S0304-4076\(01\)00080-X](https://doi.org/10.1016/S0304-4076(01)00080-X)
- Charles, V., Aparicio, J., & Zhu, J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *European Journal of Operational Research*, 279(3), 929–940. <https://doi.org/10.1016/j.ejor.2019.06.025>

- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444. [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Cherchye, L., Moesen, W., Rogge, N., & van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research*, 82(1), 111–145. <https://doi.org/10.1007/s11205-006-9029-7>
- Cordero, J. M., Polo, C., Santín, D., & Simancas, R. (2018). Efficiency measurement and cross-country differences among schools: A robust conditional nonparametric analysis. *Economic Modelling*, 74, 45–60.
- Cordero, J. M., Salinas-Jiménez, J., & Salinas-Jiménez, M. M. (2017). Exploring factors affecting the level of happiness across countries: A conditional robust nonparametric frontier analysis. *European Journal of Operational Research*, 256(2), 663–672. <https://doi.org/10.1016/j.ejor.2016.07.025>
- Cordero, J. M., Santín, D., & Simancas, R. (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of the Operational Research Society*, 68, 364–376.
- Corrente, S., Figueira, J. R., & Greco, S. (2021). Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. *European Journal of Operational Research*, 291(2), 738–756. <https://doi.org/10.1016/j.ejor.2020.09.036>
- Council of the European Union (2011). council recommendation of 28 June 2011: 'Youth on the move' - promoting the learning mobility of young people. *Official Journal of the European Union*, C(199), 1–5.
- Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis* (Vol. 4). Boston, MA: Springer. <https://doi.org/10.1007/978-0-387-35231-2>
- De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363. <https://doi.org/10.1057/jors.2015.92>
- De Witte, K., Rogge, N., Cherchye, L., & Van Puyenbroeck, T. (2013). Economies of scope in research and teaching: A non-parametric investigation. *Omega*, 41(2), 305–314. <https://doi.org/10.1016/j.omega.2012.04.002>
- De Witte, K., & Schiltz, F. (2018). Measuring and explaining organizational effectiveness of school districts: Evidence from a robust and conditional Benefit-of-the-Doubt approach. *European Journal of Operational Research*, 267(3), 1172–1181. <https://doi.org/10.1016/j.ejor.2017.12.034>
- De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, 45(17), 2401–2412.
- De Witte, K., & Rogge, N. (2010). To publish or not to publish? On the aggregation and drivers of research performance. *Scientometrics*, 85(3), 657–680.
- De Witte, K., & Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review*, 30(4), 641–653.
- D'Inverno, G., & De Witte, K. (2020). Service level provision in municipalities: A flexible directional distance composite indicator. *European Journal of Operational Research*, 286(3), 1129–1141. <https://doi.org/10.1016/j.ejor.2020.04.012>
- European Commission (2010). *EUROPE 2020: A European strategy for smart, sustainable and inclusive growth* (Tech. Rep.). Brussels, Belgium: European Commission.
- European Commission and Directorate-General for Education, Youth, Sport and Culture (2021). *Education and training monitor 2021 - executive summary*. Publications Office of the European Union.
- European Education and Culture Executive Agency (2023). *Mobility Scoreboard: Higher education background report 2022/2023* (Tech. Rep.). Luxembourg: Publications Office of the European Union.
- Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress and efficiency change in industrialized countries. *American Economic Review*, 84(1), 66–83.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A (General)*, 120(3), 253. <https://doi.org/10.2307/2343100>
- Figueira, J. R., & Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research*, 139(2), 317–326. [https://doi.org/10.1016/S0377-2217\(01\)00370-8](https://doi.org/10.1016/S0377-2217(01)00370-8)
- Fusco, E. (2015). Enhancing non-compensatory composite indicators: A directional proposal. *European Journal of Operational Research*, 242(2), 620–630.
- Fusco, E. (2023). Potential improvements approach in composite indicators construction: The multi-directional benefit of the doubt model. *Socio-Economic Planning Sciences*, 85, 101447.
- Giambona, F., Vassallo, E., & Vassiliadis, E. (2011). Educational systems efficiency in European Union countries. *Studies in Educational Evaluation*, 37(2–3), 108–122. <https://doi.org/10.1016/j.stueduc.2011.05.001>
- Gibari, S. E., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics*, 89, 1–24. <https://doi.org/10.1007/s11573-018-0902-z>



- Giménez, V., Prior, D., & Thieme, C. (2007). Technical efficiency, managerial efficiency and objective-setting in the educational system: An international comparison. *Journal of the Operational Research Society*, 58(8), 996–1007. <https://doi.org/10.1057/palgrave.jors.2602213>
- Giménez, V., Thieme, C., Prior, D., & Tortosa-Ausina, E. (2017). An international comparison of educational systems: A temporal analysis in presence of bad outputs. *Journal of Productivity Analysis*, 47(1), 83–101. <https://doi.org/10.1007/s1123-017-0491-9>
- Giménez, V., Thieme, C., Prior, D., & Tortosa-Ausina, E. (2019). Comparing the performance of national educational systems: Inequality versus achievement? *Social Indicators Research*, 141(2), 581–609. <https://doi.org/10.1007/s11205-018-1855-x>
- Greco, S., Ehrgott, M., & Figueira, J. R. (2016). *Multiple criteria decision analysis* (Vol. 233). New York: Springer. <https://doi.org/10.1007/978-1-4939-3094-4>
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141, 61–94.
- Henriques, A. A., Fontes, M., Camanho, A. S., D’Inverno, G., Amorim, P., & Silva, J. G. (2022). Performance evaluation of problematic samples: A robust nonparametric approach for wastewater treatment plants. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04629-z>
- Kao, C., & Hung, H.-T. (2003). Ranking university libraries with a posteriori weights. *Libri*. <https://doi.org/10.1515/LIBR.2003.282>
- Karagiannis, G., & Paschalidou, G. (2017). Assessing research effectiveness: A comparison of alternative nonparametric models. *Journal of the Operational Research Society*, 68(4), 456–468. <https://doi.org/10.1057/s41274-016-0168-1>
- Li, M., & Ai, N. (2022). The covid-19 pandemic: The watershed moment for student mobility in Chinese universities? *Higher Education Quarterly*, 76(2), 247–259.
- Li, Q., & Racine, J. S. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 16, 485–512.
- Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadística*, 4(2), 209–242. <https://doi.org/10.1007/BF03006863>
- McMeekin, N., Wu, O., Germeni, E., & Briggs, A. (2020). How methodological frameworks are being developed: Evidence from a scoping review. *BMC Medical Research Methodology*, 20(1), 173. <https://doi.org/10.1186/s12874-020-01061-4>
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., & Giovannini, E. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Organisation for Economic Co-operation and Development.
- Nuzulismah, R. S., Azis, A., Sensuse, D. I., Kautsarina, & Suryono, R. R. (2021). Success factors & challenges for mobile collaborative learning implementation in higher education. *2021 international conference on advanced computer science and information systems (icacsis)* (pp. 1–9). IEEE. <https://ieeexplore.ieee.org/document/9631361/>
- Oliveira, R., Zanella, A., & Camanho, A. S. (2019). The assessment of corporate social responsibility: The construction of an industry ranking and identification of potential for improvement. *European Journal of Operational Research*, 278(2), 498–513. <https://doi.org/10.1016/j.ejor.2018.11.042>
- Oliveira, R., Zanella, A., & Camanho, A. S. (2020). A temporal progressive analysis of the social performance of mining firms based on a Malmquist index estimated with a Benefit-of-the-Doubt directional model. *Journal of Cleaner Production*, 267, 121807. <https://doi.org/10.1016/j.jclepro.2020.121807>
- Pastor, J. T., & Lovell, C. A. K. (2005). A global Malmquist productivity index. *Economics Letters*, 88(2), 266–271. <https://doi.org/10.1016/j.econlet.2005.02.013>
- Pereira, A. A., & Pereira, M. A. (2023). Energy storage strategy analysis based on the Choquet multi-criteria preference aggregation model: The Portuguese case. *Socio-Economic Planning Sciences*, 85, 101437. <https://doi.org/10.1016/j.seps.2022.101437>
- Pereira, M. A., Camanho, A. S., Figueira, J. R., & Marques, R. C. (2021). Incorporating preference information in a range directional composite indicator: The case of Portuguese public hospitals. *European Journal of Operational Research*, 294(2), 633–650. <https://doi.org/10.1016/j.ejor.2021.01.045>
- Pereira, M. A., Figueira, J. R., & Marques, R. C. (2020). Using a Choquet integral-based approach for incorporating decision-maker’s preference judgments in a Data Envelopment Analysis model. *European Journal of Operational Research*, 284(3), 1016–1030. <https://doi.org/10.1016/j.ejor.2020.01.037>
- Pereira, M. A., Machete, I. F., Ferreira, D. C., & Marques, R. C. (2020). Using multi-criteria decision analysis to rank European health systems: The Beveridgean financing case. *Socio-Economic Planning Sciences*, 72, 100913. <https://doi.org/10.1016/j.seps.2020.100913>
- Pereira, M. A., & Marques, R. C. (2022). Is sunshine regulation the new prescription to brighten up public hospitals in Portugal? *Socio-Economic Planning Sciences*. <https://doi.org/10.1016/j.seps.2021.101219>

- Pereira, M. A., & Marques, R. C. (2022). The ‘Sustainable Public Health Index’: What if public health and sustainable development are compatible? *World Development*, 149, 105708. <https://doi.org/10.1016/j.worlddev.2021.105708>
- Pereira, M. A., Vilarinho, H., D’Inverno, G., & Camanho, A. S. (2023). A regulatory robust conditional approach to measuring the efficiency of wholesale water supply and wastewater treatment services. *Utilities Policy*, 83, 101594. <https://doi.org/10.1016/j.jup.2023.101594>
- Rogge, N. (2011). Granting teachers the “benefit of the doubt” in performance evaluations. *International Journal of Educational Management*, 25(6), 590–614.
- Rogge, N. (2018). Composite indicators as generalized benefit-of-the-doubt weighted averages. *European Journal of Operational Research*, 267(1), 381–392.
- Rogge, N. (2018). On aggregating benefit of the doubt composite indicators. *European Journal of Operational Research*, 264(1), 364–369.
- Rogge, N., De Jaeger, S., & Lavigne, C. (2017). Waste performance of NUTS 2-regions in the EU: A conditional directional distance Benefit-of-the-Doubt model. *Ecological Economics*, 139, 19–32. <https://doi.org/10.1016/j.ecolecon.2017.03.021>
- Sala-Garrido, R., Mocholí-Arce, M., & Molinos-Senante, M. (2021). Assessing the quality of service of water companies: A ‘Benefit of the Doubt’ composite indicator. *Social Indicators Research*, 155(1), 371–387. <https://doi.org/10.1007/s11205-020-02588-1>
- Sarrico, C. S., & Dyson, R. G. (2004). Restricting virtual weights in data envelopment analysis. *European Journal of Operational Research*, 159(1), 17–34. [https://doi.org/10.1016/S0377-2217\(03\)00402-8](https://doi.org/10.1016/S0377-2217(03)00402-8)
- Silva, M. C., Camanho, A. S., & Barbosa, F. (2020). Benchmarking of secondary schools based on Students’ results in higher education. *Omega*, 95, 102119. <https://doi.org/10.1016/j.omega.2019.102119>
- Simos, J. (1990). *Evaluer l’impact sur l’environnement: Une approche originale par l’analyse multicritère et la négociation*. Lausanne: Presses Polytechniques et Universitaires Romandes.
- Simos, J. (1990). *L’évaluation environnementale: Un processus cognitif négocié (Unpublished doctoral dissertation)*. Suisse: EPFL-Lausanne.
- 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(2), 577–608. <https://doi.org/10.1007/s10479-019-03329-5>
- Szuwarzyński, A. (2019). Benefit of the doubt approach to assessing the research performance of Australian universities. *Higher Education Quarterly*, 73(2), 235–250. <https://doi.org/10.1111/hequ.12184>
- Szuwarzyński, A. (2022). Efficiency evaluation of graduation process in Australian public universities. *Economic Research-Ekonomska Istraživanja*, 35(1), 4220–4236. <https://doi.org/10.1080/1331677X.2021.2013267>
- Tauchmann, H. (2012). Partial frontier efficiency analysis. *The Stata Journal*, 12(3), 461–478.
- Van Puyenbroeck, T. (2018). On the output orientation of the Benefit-of-the-Doubt-Model. *Social Indicators Research*, 139(2), 415–431. <https://doi.org/10.1007/s11205-017-1734-x>
- Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with benefit-of-the-doubt weights. *European Journal of Operational Research*, 256(3), 1004–1014.
- Verschelde, M., & Rogge, N. (2012). An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional Data Envelopment Analysis approach. *European Journal of Operational Research*, 223(1), 214–225. <https://doi.org/10.1016/j.ejor.2012.05.044>
- Vilarinho, H., D’Inverno, G., Nóvoa, H., & Camanho, A. S. (2023). Performance analytics for regulation in retail water utilities: Guiding asset management by identifying peers and targets. *Utilities Policy*, 82, 101559. <https://doi.org/10.1016/j.jup.2023.101559>
- Walheer, B. (2024). A sequential benefit-of-the-doubt composite indicator. *European Journal of Operational Research*, 316, 228–239. <https://doi.org/10.1016/j.ejor.2024.01.029>
- Zanella, A., Camanho, A. S., & Dias, T. G. (2013). Benchmarking countries’ environmental performance. *Journal of the Operational Research Society*, 64(3), 426–438. <https://doi.org/10.1057/jors.2012.62>
- Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research*, 245(2), 517–530. <https://doi.org/10.1016/j.ejor.2015.03.036>