VILNIUS UNIVERSITY

Dovilė
STUMBRIENĖ

Data Envelopment Analysis for Effectiveness Analysis of Education Systems

SUMMARY OF DOCTORAL DISSERTATION

Natural Sciences,
Informatics N 009

VILNIUS 2019

This dissertation was written between 2014 and 2018 at Vilnius University.

Academic supervisor – Prof. Dr. Audronė, Jakaitienė (Vilnius University, Natural Sciences, Informatics – N 009).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Olga Kurasova (Vilnius University, Natural Sciences, Informatics – N 009).

Members:
Prof. Dr. Tommaso Agasisti (Politecnico di Milano School of Management, Italy, Social Sciences, Economy – S 004),
Prof. Habil. Dr. Juozas Augutis (Vytautas Magnus University, Natural Sciences, Informatics – N 009),
Prof. Dr. Tomas Baležentis (Vilnius University, Social Sciences, Economy – S 004),
Dr. Remigijus Paulavičius (Vilnius University, Natural Sciences, Informatics – N 009).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 1 p. m. on 3rd July 2019 in auditorium 203 of the Institute of Data Science and Digital Technologies of Vilnius University. Address: Akademijos St. 4, LT-08412 Vilnius, Lithuania.

The summary of the doctoral dissertacion was distributed on the 3rd of June, 2019.

The text of this dissertation can be accessed at the libraries of Vilnius University, as well as on the website of Vilnius University: www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

Dovilė
STUMBRIENĖ

Duomenų apgaubties metodas švietimo sistemų efektyvumo analizėje

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,
Informatika N 009

VILNIUS 2019

Disertacija rengta 2014–2018 metais Vilniaus universitete.

Mokslinė vadovė – prof. dr. Audronė Jakaitienė (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Gynimo taryba:

Pirmininkė – prof. dr. Olga Kurasova (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Nariai:
prof. dr. Tommaso Agasisti (Politecnico di Milano verslo mokykla, Italija, socialiniai mokslai, ekonomika – S 004),
prof. habil. dr. Juozas Augutis (Vytauto Didžiojo universitetas, gamtos mokslai, informatika – N 009),
prof. dr. Tomas Baležentis (Vilniaus universitetas, socialiniai mokslai, ekonomika – S 004),
dr. Remigijus Paulavičius (Vilniaus universitetas, gamtos mokslai, informatika – N 009).

Disertacija ginama viešame Gynimo tarybos posėdyje 2019 m. liepos 3 d. 13 val. Vilniaus universiteto Duomenų mokslo ir skaitmeninių technologijų instituto 203 auditorijoje. Adresas: Akademijos g. 4, LT-08412 Vilnius, Lietuva.

Disertacijos santrauka išsiuntinėta 2019 m. birželio mėn. 3 d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: https://www.vu.lt/naujienos/ivykiu-kalendorius

1 Introduction

1.1 Research area

Education is one of the milestones that guarantee the well-being of nations in the long run. Education is a multidimensional system consisting of subsystems (pre-primary, primary and secondary education, vocational training, tertiary education, lifelong learning and etc.) and the status of education is determined by many different factors. The complexity of the system creates space for different types of tasks, especially when it comes to assessing and comparing the entire education systems between countries. The comparison of those countries' education systems is a difficult multi-criteria optimization task.

Effectiveness of education systems can be defined as the degree to which an education system achieve desired goals and effects [17]. In the context of education systems in European countries, goals and effects could be represented in terms of education systems achievement according the strategic framework "Education and Training 2020" (ET2020), an education system that contributes to greater levels of these achievement is considered more effective than another education system.

Whilst it is possible to evaluate effectiveness using individual sub-indicators, it is not a trivial task to conduct multi-dimensional evaluations. The construction of Composite Indicators (CIs) might be considered when taking into account several sub-indicators simultaneously. A CI incorporates several indicators into a single summary measure that can give a notion about the status of a system. The usual way to aggregate sub-indicators is Simple Additive Weighting (SAW) with equal weights. However, due to the different environment and conditions, European countries are diverse with respect to education systems. Therefore more sophisticated methods should be implemented for the assessment of performance of education systems. Taking into account that there is no single way to improve the education systems in all European countries; it is important to estimate the performance of each country and to provide guidelines that particular country should follow to improve her performance of education system. As pointed out Silva et al. (2017) [49] measuring performance in absolute terms is often less valuable than making comparisons with other countries, and provide examples of good education practices that under-performing countries should follow to improve the performance of their education systems.

The main object of this study is Data Envelopment Analysis (DEA) and its implementation for the effectiveness estimation of education systems in 29 European countries

1.2 Relevance of the Research

The European Commission monitors the performance of education systems in Member States according to the strategic framework ET2020. However, the importance of the performance of education systems at country level is still underestimated – only a small number of frontier-based efficiency studies of the education sector have focused on country level or cross-country analyses (Decision Making Unit (DMU) is a country), meanwhile the evaluation of schools (DMU is a school or university) is a widely researched topic [59].

There are several papers ([3–6, 23, 34, 36, 51, 53, 55]) that focus on cross-country comparisons (DMU is a country). This topic is especially urgent because the enlargement of the EU in 2004–2013; the new EU countries do not make one type of education system and moved in different directions [14, 63].

However, a small number of studies consider the system of a country as a unit of assessment where sub-indicators represent all levels of its education system. Most of the studies on education efficiency using country level data focused exclusively on a single educational stage. However, the comparison among educational systems based on primary, secondary or tertiary education alone does not represent the overall education system. Only analyses involving all educational levels together can accurately represent the education system of a country. To the best of our knowledge, only the paper of Bogetoft et al. (2015) [15] used data covering all educational stages (primary, secondary and tertiary). This implies that the analysis of the education system as a whole at country level can be considered as state-of-the art in education research.

In recent years, the DEA method has been increasingly used to aggregate sub-indicators [18, 22, 37, 42, 43, 61, 62], but the SAW method is also applicable due to its ease of use and transparency. The DEA allows flexible weights, as the SAW - fixed weights only. No one has proposed the method which allows transition from fixed to flexible weights while calculating a CI.

The original DEA model allows total flexibility of the weights, i.e. each DMU maximizes its efficiency score, given the inputs consumed and the outputs attained. The flexibility of the weights' selection allows some sub-indicators to be assigned a zero weight. Due to full flexibility, many DMUs will be able to achieve the maximum DEA efficiency score [41]. DEA loses discrimination power when the number of sub-indicators increase compared with the number of the DMUs, this is undesirable when countries ranking is carried out.

The ratio between number of sub-indicators and DMUs are described in academic literature [12, 16, 29, 30, 50]. However, even if the ratio is satisfied, it does not guarantee desirable discriminatory power. In the scientific literature the ratios between the sub-indicators to be measured and the number of DMUs analysed are not described when calculating the CI and performing DMU rankings. For this type of task, where only output sub-indicators are available and input indicators are dummy, the maximum discrimination power of the model is required in order to perform country ranking.

In the DEA the restriction of weights is the most often used to improve discrimination power [28, 30, 48, 54, 56, 57, 60, 62]. To the extent of our knowledge, weight flexibility has not been proposed so far, allowing a gradual transition from fixed to flexible weighting in the DEA model and ensuring a sufficient discriminatory power of DEA model as well as inclusion of all sub-indicators.

When the number of available sub-indicators, compared to the number of DMUs being evaluated, will increase dramatically, an attractive alternative to address the discrimination problem of DEA model will be to reduce the data dimension before applying DEA. The papers [1, 2, 58] showed that PCA can help improving discrimination power in DEA, which often fails when there are too many sub-indicators in relation to the number of DMUs, and give more reliable efficiency measurement in small samples. However the idea to combine PCA and DEA is rarely applied. The use of PCA-DEA approach for performance assessment can be found in different fields: aviation [1], manufacturing [11, 46], logistics [9, 20, 38], ecology [40, 44], agriculture [27], finance [39] and health [10]. To our knowledge, there are two papers [2, 26] where the hybrid PCA-DEA approach was applied for education data (assessed performance of university departments and schools).

1.3 Objective and tasks of the research

The objective of this study is to propose a methodology for evaluating the effectiveness of education systems based on the DEA analysis.

The following tasks were identified:

1. To investigate the literature associated with the education efficiency research with a special focus on the studies with country level analyses.

2. To aggregate selected sub-indicators of the education system applying the CI methodology.

3. To apply the DEA model for the evaluation of the performance of education at system level.

4. To propose new type weights restrictions in the DEA model for the implementation the transition from fixed to flexible weighting systems.

5. To assess the suitability of data dimension reduction method for the assurance sufficient discrimination power for the DEA model.

6. To propose the methodology for performance evaluation of education systems according to the research of this study.

1.4 Scientific novelty

CIs have been accepted as a useful tool for conducting performance assessment and rankings calculations in various fields, but only a few CIs were constructed in the area of education. The Lithuanian education system was not evaluated in this context. This study proposes a new CI to summarise the performance of European countries' education systems considering all educational stages (pre-primary, primary, secondary, tertiary education and life long learning) and compare the performance at country level in the light of the Europe 2020 strategy.

In calculating the CIs of European education systems, we have shown that the ratio between sub-indicators and the number of countries assessed is not sufficient when calculating a CI for country ranking. The novelty of this study in the field of informatics is the proposal of the gradual transition from fixed to flexible weighting in the DEA model, which solves the problem of insufficient discrimination of the DEA model, insures the inclusion of all sub-indicators in the CI and provides additional information that may be used to improve system performance. Moreover, the use of the hybrid PCA-DEA model as an alternative to weight restrictions has been empirically investigated.

The practical novelty of this work is the calculation of a new CI for assessing the effectiveness of European education systems and the proposed methodology for analysing the effectiveness of education systems. This work is important not only for the development of educational research theory, but also for practical applicability. Over the last decade, new strategies for the development of European education systems have been implemented. The EU has adopted targets that should be achieved in five areas by 2020, including education. Based on the methodology proposed for the assessment of the effectiveness of education systems in this work, and taking into account the objectives of ET2020, one can assess areas for improvement in European education systems. In addition, the proposed methodology can be a useful tool for setting future targets not only in education but also in other areas.

1.5 Statements to be defended

1. Evaluation of education systems' effectiveness and calculation of CI should be performed using the DEA instead of the traditional SAW.

2. Weight restrictions must be introduce in the DEA when the number of sub-indicators is disproportionate to the number of countries evaluated.

3. The DEA model with ARI weight restrictions can be used to construct the CIs implementing transition from fixed to flexible weighting systems.

4. The DEA model with ARI weight restrictions and a new flexibility parameter $\sigma \in [0; 1]$ allows gradual transition from fixed to flexible weighting systems.

5. The discrimination power of PCA-DEA approach, as an alternative to weight restrictions, is higher than the DEA model when assessing the effectiveness of European education systems.

1.6 Outline of the thesis

The dissertation consists of seven Chapters, References and an Appendix. The chapters of the dissertation are as follows: Introduction; Composite indicator for effectiveness evaluation of education systems; DEA model for effectiveness evaluation; Weights restrictions in the DEA model; Analysis of data by means of PCA-DEA; Methodology for effectiveness estimation of education systems; The main resuls and conclusions.

2 Composite indicator for effectiveness evaluation of education systems

The evaluation of effectiveness of education systems has become a priority in education policy agendas internationally. Whilst it is possible to evaluate effectiveness using individual indicators and assess their evolution over time, obtaining an overview of the effectiveness of an education system and comparing it with other systems is not a trivial task. As noted by Grupp and Mogee (2004) [35], individual indicators describe various aspects of multi-dimensional processes, but they do not measure processes as a whole. To be able to evaluate the implication of policies and develop a culture of improvement based on reliable information, it is necessary to aggregate data on the performance of individual processes to obtain a stylised and simplified view of complex systems. CIs provide the key contribution towards the achievement of this objective.

A CI incorporates several indicators into a single summary measure that gives a notion about the status of a system. According to Saltelli (2007) [47], CIs provide a big picture of multidimensional processes, which can be extremely useful to guide public policy discussions and attract public interest. In particular, CIs can be used to monitor the impact of national policies on education systems and contribute to continuous improvement.

In the education context, in 2006 the Canadian Council on Learning created the world's first Composite Learning Index (CLI) with the purpose of measuring progress in lifelong learning over time. In 2010, this approach was adopted for the country-level assessment of lifelong learning by the German Bertelsmann foundation. The composite indicator for the country-level assessment of lifelong learning in the EU Member States was named the European Lifelong Learning Indicators (EELI) index. In 2012, the German Learning Atlas was created, the first indicator-based regional monitoring instrument for lifelong learning in Europe [52]. It allows the observation and comparison of the conditions for lifelong learning in all 412 German administrative districts and cities, as well as in the federal states.

According to the handbook of OECD [24], the construction of CI implies several stages. These include the selection of sub-indicators, an exploratory analysis of sub-indicators and treatment of data, the application of a normalization process, the specification of the weights for the sub-indicators and the selection of the functional form of the aggregator function.

2.1 Sub-indicators used for the evaluation of countries' education systems

The strategic framework for European cooperation in education and training (ET2020) has set four common EU objectives to address challenges in education and training by 2020. These are as follows: (1) Making lifelong learning and mobility a reality; (2) Improving the quality and efficiency of education and training; (3) Promoting equity, social cohesion, and active citizenship; (4) Enhancing creativity and innovation, including entrepreneurship, at all levels of education and training. To achieve these objectives, a set of targets have also been stated:

1. Reducing the rate of early leavers from education and training aged 18-24 below 10%.
2. At least 40% of people aged 30-34 should have received some form of higher education.
3. At least 95% of children (from 4 to compulsory school age) should participate in early childhood education.
4. The share of employed graduates (aged 20-34 with at least upper secondary education attainment and having left education 1-3 years ago) should be at least 82%.
5. At least 15% of adults should participate in learning.

6. Less than 15% of 15-year-olds should be under-skilled in reading, mathematics and science.

7. At least 20% of higher education graduates and 6% of 18-34 year-olds with an initial vocational qualification should have spent some time studying or training abroad.

These targets cover several different aspects of education and training, namely: (1) Early leavers, (2) Tertiary education attainment, (3) Early childhood education and care, (4) Employment rate of recent graduates, (5) Adult participation in learning, (6) Under-achievement in reading, maths and science, (7) Learning mobility. Of the seven targets defined, only the target on learning mobility is still awaiting the required compilation of cross-national data [31]. For the first five dimensions of the ET2020 strategy, data is available in the Eurostat database. For dimension 6, associated with low achievement, we used the PISA (Programme for International Student Assessment) dataset, available from OECD.

For the assessment of education systems, we have considered indicators covering the first six dimensions of the ET2020 strategy, and also added two more indicators. The first represents Top achievement in reading, maths and science, to enable an enhanced evaluation of students' literacy skills. The second is Upper secondary or tertiary education attainment at the age of 25-64, in order to evaluate the likelihood of having the minimum necessary qualifications to actively participate in social and economic life in the 21st century.

Table 1: Sub-indicators for the construction of CIs.

	Sub-indic.	Description and source
Y_1^*	Early leavers from education and training	The percentage of the population aged 18-24 with at most lower secondary education and who were not in further education or training during the last 4 weeks [Eurostat]
Y_2	Tertiary educational attainment	The share of population aged 30-34 years who have successfully completed tertiary-level education [Eurostat]
Y_3	Early childhood education	The share of population aged 4 to the age when the compulsory education starts who are participating in early education [Eurostat]
Y_4	Employment rates of recent graduates	The share of employed graduates (20-34 years) having left education and training 1-3 years before the reference year [Eurostat]
Y_5	Lifelong learning	The participation rate of adults (25-64 years) in education and training [Eurostat]
Y_6^*	PISA Low achievers	The percentage of PISA Low achievers (below Level 2) in reading, maths and science [OECD]
Y_7	PISA Top achievers	The percentage of PISA Top achievers (Level 5 or 6) in reading, maths and science [OECD]
Y_8	Upper secondary educational attainment	The percentage of people aged 25-64 who have successfully completed at least upper secondary education [Eurostat]

Sub-indicators $Y1_1^*$ and Y_6^* are measured in such a way that lower values represent better performance.

The CI for education systems evaluation was estimated for 29 European countries, using data from Eurostat [1] and OECD [2] databases for year 2015. Table 1 provides a summary of the sub-indicators used in this study.

2.2 Data treatment

The first step of the data treatment consisted of adjusting the scale of the indicators so that higher values of all sub-indicators correspond to better performance. The values of sub-indicators (Y_1^* and Y_6^*) were converted using the complement to 100 %. Table 2 provides a descriptive statistics of analysed sub-indicators for 29 European countries for year 2015 and ET2020 targets.

Another step required for the construction of many CIs is the normalisation of data. In the comparison of the results of the SAW model and the DEA model with equal pure weights, the original sub-indicators were previously normalised as shown in expression:

$$Y^{norm} = \frac{Y - \bar{Y}}{SD_Y} + \left| min\left(\frac{Y - \bar{Y}}{SD_Y}\right) \right| + \epsilon, \epsilon = 0.001. \qquad (1)$$

This ensures that all values of the normalised dataset are strictly positive.

[1] http://ec.europa.eu/eurostat/web/education-and-training/data/database

[2] http://www.oecd.org/pisa/data/2015database/

Table 2: Descriptive statistics of analysed sub-indicators for 29 European countries and ET2020 targets.

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
ET2020 targets	90.0	40.0	95.0	82.0	15.0	85.0	-	-
Mean	90.2	41.4	92.2	76.3	12.6	78.5	8.3	80.2
SD	4.3	8.9	7.0	11.2	9.3	7.1	3.1	11.0
Median	90.8	43.4	95.0	79.5	9.7	80.2	8.8	82.7
Max	97.2	57.6	100.0	92.0	31.3	89.8	13.2	93.5
Min	80.0	25.3	73.8	45.2	1.3	59.5	2.0	45.1

2.3 Specification of aggregation functions and weight restrictions

According to OECD handbook [24], the main problem in the construction of CIs concerns the aggregation of the information. This involves the specifications of weights and definition of the functional form for the aggregation. In this paper, we will discuss different modelling alternatives proposed in the literature by dividing the specification of weights into two categories: fixed and flexible weighting systems.

Fixed weighting system

Let n be the number of countries whose CIs are to be calculated based on s sub-indicators. Let y_{rj} denote the value of the output sub-indicator r ($r = 1, \ldots, s$) observed for country j ($j = 1, \ldots, n$). All sub-indicators are measured in a scale indicating that higher values of the sub-indicator represent better performance (i. e., y_{rj} are desirable outputs). The purpose of the construction of a composite indicator CI_j for unit j is to aggregate individual sub-indicators y_{rj} into a single summary measure of performance.

The SAW method involves normalising the data and then attaching weights to the sub-indictors to produce a CI score, as shown in expression:

$$CI_j^S = \sum_{r=1}^{s} w_r y_{rj}. \tag{2}$$

In expression 2 $w_r \in \mathbb{R}^+$ is the weight attached to each sub-indicator y_{rj} in the assessment of performance of unit j. In applications of the SAW method, the weights assigned to the indicators are often identical, which implies that all indicators have the same impact on performance (Freudenberg, 2003). In such cases, the values of the individual weights become $w_r = \frac{1}{s}$ and the sum of all weights is equal to one ($\sum_{r=1}^{s} w_r = 1$). Consequently, the CI represents a simple weighted arithmetic average of the indicators. The CI score can also be used to rank the performance of the units under assessment, with higher values corresponding to better performance. The results (CIs and Ranks) of the SAW model 2 with equal importance attributed to all indicators are presented in Table 3 (collums CI_{abs}^S and R_{abs}^S).

Flexible weighting system

Flexible weighting systems can be implemented using DEA models, under the BoD approach, a term coined by Cherchye et al. [21]. It suggests the use of a dummy input equal to 1 and multiple outputs that represent individual performance indicators to be combined in the composite measure. As Cherchye et al. [21] pointed out, the BoD model is formally tantamount to the original input-oriented DEA CRS model [19]. In the absence of reliable and consensual information about the weights to be used in the aggregation stage, this method endogenously selects those weights that maximise the CI score for the unit under assessment. Thus, each unit can be assessed with its own weights, emphasising aspects with good performance.

In this research we adopt the alternative formulation of BoD CIs proposed by Zanella et al. [61], also with a single dummy input, but with an output-oriented CRS formulation, as shown in 3. It is more consistent with the goal of the effectiveness assessment (i.e., maximization of outcomes) and allows a direct estimation of the targets on the frontier of the production possibility set. It also enables an easier implementation of weight restrictions, bridging the gap between the flexible weighting strategies (operationalised using model 3) and fixed weighting strategies (operationalised using expression 2).

$$
\begin{aligned}
\frac{1}{CI_{jO}^{D}} &= \min \nu, \\
\sum_{r=1}^{s} u_r y_{rjO} &= 1, \\
\sum_{r=1}^{s} u_r y_{rj} - \nu &\leqslant 0, \\
u_r \geqslant 0, \nu &\geqslant 0.
\end{aligned} \tag{3}
$$

Model 3 uses linear programming to estimate a CI for each DMU (country) in terms of the achievements in different dimensions, corresponding to the outputs specified. The estimation of the CI involves solving one model for each DMU j_O under assessment, thus enabling the estimation of weights that are DMU-specific. The CI for DMU j_O is given by $\frac{1}{\nu^*}$, where the symbol * signals the optimal solution to the linear programming model 3. The value of the CI for DMU j_O ranges between 0 (worst) and 1 (best). y_{rj} is the rth output of the jth DMU. u_r and ν are the pure weights given to the rth output ($r = 1, \ldots, s$) and dummy input, respectively, in the estimation of the CI score for DMU j_O. The optimal virtual output weights are given by the product of the outputs observed at the DMU under assessment by the respective pure weights, $u_r y_{rjO}$ ($r = 1, \ldots, s$).

For the identification of peers and targets, the dual formulation (or "envelopment formulation") of the DEA model 3 can be used, as shown:

$$
\begin{aligned}
\max\theta, \\
\theta y_{rjO} - \sum_{j=1}^{n} \lambda_j y_{rj} \leqslant 0, \\
\sum_{j=1}^{n} \lambda_j \leqslant 1, \\
\lambda_j \geqslant 0.
\end{aligned}
\tag{4}
$$

For each assessed DMU j_O, the solution of the model 5 seeks to identify a comparator, i. e., a composite DMU corresponding to a linear combination of efficient DMUs (peers), whose outputs $\sum_{j=1}^{n} \lambda_j y_{rj}$, $r = 1, \ldots, s$ dominate the output levels y_{rjO} $(r = 1, \ldots, s)$ of the DMU j_O under assessment. This linear combination of the efficient DMUs defines the frontier of the production possibility set determined by the DEA model. The decision variables of the LP model 5 are λ_j $(j = 1, \ldots, n)$. If $\lambda_j > 0$, then the corresponding DMU j is peer to DMU j_O under assessment.

$$
y^*_{rj0} = \sum_{j=1}^{n} \lambda^*_j y_{rj}, r = 1, \ldots, s. \tag{5}
$$

The targets $(y^*_{rjO}$, $r = 1, \ldots, s)$ corresponding to efficient operation for DMU j_O are obtained as shown in 5, where λ_j^*: corresponds to the optimal solution obtained for variables λ_j $(j = 1, \ldots, n)$ using the model 4.

The standard DEA technique in this study showed limited discrimination power. The results (CIs and Ranks) of the DEA model 3 are presented in Table 4 (collumns CI_0^D and R_0^D). More than two-thirds of the countries obtained a CI score equal to 1. The obtained results show that only 9 countries out of 29 are not Pareto-optimal. In the case when Pareto optimal points belong to the convex hul of the data set, the application of the standard DEA technique, defines the unit efficiency of all Pareto optimal DMUs [25]. In this study the discrimination of the DEA model have been increased through the use of weigthts restrictions in the DEA model and the dimension reduction of sub-indicators before running the DEA model.

3 Weight restrictions in the DEA model

The original DEA model [19] allows total flexibility in the selection of the weights to be attached to the inputs and outputs. The flexibility in the choice of weights, which is the strength of a DEA analysis, may also be a weakness, as it allows some sub-indicators to be assigned a zero weight. Consequently, classical DEA models often identify too many DMUs as efficient due to the full flexibility allowed in the selection of weights [41]. In the context of the construction of CIs, we expect high discrimination power to enable the construction of rankings, such that having a large number of DMUs reaching an efficiency score equal to one is undesirable. In addition, having weights u_r equal to 0 is undesirable, as it means that the corresponding sub-indicator y_{rj} is ignored in the estimation of the CI score.

An important procedure to enhance discrimination and avoid the occurrence of zero weights is the use of weight restrictions. Weight restrictions can be imposed using different formulations, which can be classified as restrictions imposed to the pure weights and restrictions imposed to virtual weights (or sub-indicator shares). Restrictions imposed to pure weights are dependent on the units of measurement of the original sub-indicators.

3.1 Restrictions to the pure weights

The most common restrictions imposed directly to pure weights are Absolute Weight Restrictions [28] and Assurance Regions [57]. Absolute weight restrictions are usually introduced to prevent the inputs or outputs from being over emphasised or ignored in the assessment. However, they have several limitations. They may render infeasible solutions or lead to the underestimation of the relative efficiency scores [8, 45]. Another difficulty associated with absolute weight restrictions is the interpretation of the meaning of bounds.

Weight restrictions in the form of Assurance Regions differ from absolute weight restrictions because instead of requiring the weights to be within certain limits, they require ratios between weights to be within certain limits. In general, weights are meaningful only on a relative basis [32], representing marginal rates of substitution between the inputs or outputs. The most prevalent type of weight restrictions used in DEA applications are assurance regions type I (ARI). As pointed out by Allen et al. [8] and Sarrico and Dyson [48], a disadvantage of this type of weight restrictions is that they are sensitive to the units of measurement of inputs and outputs. Furthermore, it is often difficult to specify meaningful marginal rates of substitution between the variables.

3.2 Restrictions to the virtual weights

Virtual weight restrictions [60] are expressed in percentual terms, and so are independent of the units of measurement of the indicators. As pointed out by Thanassoulis et al. [54], the restrictions to virtual weights tend to be computationally expensive and may lead to infeasible solutions when the bounds are loosely specified. As suggested by Wong and Beasley [60], an alternative to overcome the infeasibility problems and the computational difficulties, is to apply the above restrictions only to the virtual outputs of the DMU j_O under assessment. However, the restrictions imposed only to the DMU under assessment also have drawbacks. According to Dyson et al. [30], if the restrictions are imposed only on the virtual outputs of the DMU under assessment, they compromise the symmetry of the model with respect to all DMUs, as each DMU is assessed based on a different feasible region. Sarrico and Dyson [48] added that these restrictions imposed only on the DMU under assessment might impose unreasonable restrictions on the virtual weights of the other DMUs.

To overcome these limitations, Zanella et al. [62] proposed a new type of weight restrictions, defined as ARI restrictions, but that enable expressing the relative importance of the output indicators in percentual terms instead of specifying marginal rates of substitution. This requires the use of an "artificial" DMU representing the average values of the outputs in the sample analyzed. This type of formulation for the weight restrictions recurring to the use of an "artificial" DMU was originally proposed by Wong and Beasley [60] as a complement of DMU-specific virtual weight restrictions. If instead of restricting the virtual outputs of DMU j_O, the restrictions are imposed to the average DMU ($\bar{y}_r$), as shown in 6, all DMUs are assessed with identical restrictions. Thus, these weight restrictions in fact work as ARIs since they are no longer DMU-specific.

$$\phi_r \leqslant \frac{u_r \bar{y_r}}{\sum_{r=1}^{s} u_r \bar{y_r}} \leqslant \psi_r. \tag{6}$$

Another advantage of this type of restrictions is that the bounds become independent of the units of measurement of the outputs. The bounds ϕ_r or ψ_r of expression 6 can be interpreted as the percentual importance of output y_r in the assessment. Values of ϕ_r and ψ_r equal to 1 mean that output y_r is the only one that should be considered in the assessment, whereas values equal to 0 mean that the corresponding output should be ignored. Values ranging between 0 and 1 corresponding to varying degrees of importance are assigned to output y_r.

3.3 The comparing of the fixed and flexible weighting systems

In order to compare the fixed and flexible weighting systems, we will start by comparing the results of the fixed weighting system, with equal importance attributed to all indicators, implemented with different modelling alternatives: the SAW model and the DEA model. In the DEA framework, a structure representing equal importance of all indicators can be designed as follows:

- Restrictions with equal pure weights, applicable to normalised data:

$$u_r = u_1, r = 2, \ldots, 8. \tag{7}$$

- Restrictions with equal virtual weights (DMU-specific restrictions, applied to the country under assessment), applicable to original data:

$$u_r y_{jr0} = \frac{1}{8}, r = 1, \ldots, 8. \tag{8}$$

- ARI-type restrictions with equal virtual weights (identical restrictions applied to all countries, involving the use of a virtual DMU equal to the sample mean), applicable to original data:

$$\frac{u_r \bar{y_r}}{\sum_{r=1}^{8} u_r \bar{y_r}} = \frac{1}{8}, r = 1, \ldots, 8. \tag{9}$$

Next, we will proceed with the analysis of formulations that progressively allow further levels of weight flexibility in the DEA framework, and compare the results obtained with the fixed weighting system. The weight flexibility will be implemented through the adjustment of expression 9. This involves the use of a parameter σ, whose value, expressed in percentage, indicates the proportion of the total virtual weight $\sum_{r=1}^{s} u_r \bar{y_r}$ that is fixed:

$$\frac{u_r \bar{y_r}}{\sum_{r=1}^{s} u_r \bar{y_r}} \geq \frac{\sigma}{s}, r = 1, ..., s. \tag{10}$$

If the parameter σ is equal to 0, the resulting DEA model is equivalent to a formulation without weight restrictions (representing full flexibility of weights). If the parameter σ is equal to 1, the DEA models corresponds to a formulation with equal weights for all dimensions (representing a fixed weights scenario). Intermediate scenarios can be obtained using other values of $\sigma \in [0; 1]$, where lower values correspond to greater flexibility.

3.4 Analysis and discussion of the results obtained

Table 3 shows the results of the CI using a fixed weighting system, implemented using the SAW model with normalised data and using the DEA model with the three alternative specifications of weights (expressions 7, 8, 9). In this table $\Delta R_1 = R^D_{vir} - R^D_{abs}$ and $\Delta R_2 = R^D_{ARI} - R^D_{abs}$.

From Table 3 we can conclude that the SAW model with equal weights imposed on normalised data gives exactly the same ranking as the DEA model with equal pure weights imposed on normalised data ($R^S_{abs} = R^D_{abs}$). This confirms the possibility of using DEA models for constructing CIs as an alternative to the traditional SAW model. With appropriate normalisation and imposition of weights as formulated in the equation 7, the results of the two modelling alternatives become identical. One of the advantages of DEA is that it can avoid prior normalisation of data to conduct a relative performance assessment.

As shown in Table 3, the results obtained with alternative DEA formulations using original data are also good approximations of the DEA model with equal pure weights results. In particular, the Spearman correlation coefficient between the ranks of the DEA model with equal pure weights (R^D_{abs}) and the DEA model with equal virtual weights (R^D_{vir}) is 0.873 (p-values < 0.01). Furthermore, the Spearman correlation coefficient between the ranks of the DEA model with equal pure weights (R^D_{abs}) and the DEA model with ARI restrictions (R^D_{ARI}) is 0.942 (p-values < 0.01).

Table 3: Results for the SAW and DEA models.

Cnt.	CI^S_{abs}	R^S_{abs}	CI^D_{abs}	R^D_{abs}	CI^D_{vir}	R^D_{vir}	ΔR_1	CI^D_{ARI}	R^D_{ARI}	ΔR_2
AUT	2.638	12	0.866	12	0.825	12	0	0.818	13	1
BEL	2.506	17	0.823	17	0.676	16	-1	0.779	16	-1
BGR	1.384	26	0.454	26	0.308	28	2	0.588	27	1
HRV	1.616	25	0.531	25	0.410	26	1	0.612	26	1
CZE	2.371	19	0.779	19	0.687	15	-4	0.744	19	0
DNK	3.030	2	0.995	2	0.950	4	2	0.962	3	1
EST	2.837	7	0.932	7	0.848	11	4	0.866	8	1
FIN	2.868	6	0.942	6	0.965	3	-3	0.955	4	-2
FRA	2.629	13	0.864	13	0.877	8	-5	0.859	10	-3
DEU	2.682	11	0.881	11	0.715	14	3	0.798	14	3
GRC	1.206	28	0.396	28	0.390	27	-1	0.572	28	0
HUN	2.043	21	0.671	21	0.611	20	-1	0.691	21	0
ISL	2.483	18	0.816	18	0.881	7	-11	0.901	7	-11
IRL	2.819	8	0.926	8	0.669	17	9	0.792	15	7
ITA	1.350	27	0.443	27	0.564	22	-5	0.619	25	-2
LVA	2.320	20	0.762	20	0.557	23	3	0.700	20	0
LTU	2.601	15	0.854	15	0.593	21	6	0.749	17	2
LUX	2.627	14	0.863	14	0.863	10	-4	0.850	11	-3
NLD	2.956	5	0.971	5	0.926	6	1	0.909	6	1
NOR	3.003	4	0.986	4	0.935	5	1	0.913	5	1
POL	2.599	16	0.854	16	0.494	24	8	0.747	18	2
PRT	1.686	23	0.554	23	0.655	18	-5	0.691	22	-1
ROU	0.898	29	0.295	29	0.213	29	0	0.526	29	0
SVK	1.691	22	0.555	22	0.412	25	3	0.625	24	2
SVN	2.700	10	0.887	10	0.808	13	3	0.821	12	2
ESP	1.663	24	0.546	24	0.653	19	-5	0.684	23	-1
SWE	3.045	1	1.000	1	0.972	2	1	0.968	2	1
CHE	3.023	3	0.993	3	1.000	1	-2	1	1	-2
GBR	2.798	9	0.919	9	0.873	9	0	0.860	9	0

We used the Wilcoxon's signed-rank test for matched pairs to compare the two distributions of results, corresponding to the DEA model with equal pure weights (CI^D_{abs}) versus the DEA formulation with equal virtual weights (CI^D_{vir}), and the null hypothesis is rejected (p-value $= 0.004$). When the comparison is done between the DEA model with equal pure weights (CI^D_{abs}) versus the formulation with ARI restrictions (CI^D_{ARI}), the null hypothesis is not rejected (p-value $= 0.905$). The similarity between the average CI^D_{abs} and CI^D_{ARI} is remarkable (0.771 versus 0.779). Furthermore, only two countries (Iceland and Ireland) have change their rank by more than 3 positions between these two formulations (column ΔR_2 in Table 3).

This similarity in the results of these two modelling alternatives opens up the possibility of using DEA models with ARI weight restrictions to progressively depart from fixed weighting to flexible weighting systems. Our idea is to explore this possibility, taking advantage of some of the DEA features that are very attractive for performance management: the identification of peers and targets as by-products of the performance assessment.

Next we provide the results of the DEA model with progressive levels of flexibility allowed. The most restrictive case ($\sigma = 1$) corresponds to ARI restrictions emulating fixed weights and the least restrictive case ($\sigma = 0$) corresponds to the DEA model without weight restrictions. Intermediate scenarios corresponding to weighting systems with bounds progressively relaxed are shown in Table 4; here the difference between the most and the least

Table 4: Results for DEA model with different degrees of flexibility.

Cnt.	CI_1^D	R_1^D	$CI_{0.9}^D$	$R_{0.9}^D$	$CI_{0.5}^D$	$R_{0.5}^D$	$CI_{0.1}^D$	$R_{0.1}^D$	CI_0^D	R_0^D	ΔCI
AUT	0.818	13	0.843	12	0.922	15	0.985	19	1	1	0.182
BEL	0.779	16	0.809	16	0.921	16	0.983	20	0.995	21	0.216
BGR	0.588	27	0.624	27	0.768	27	0.913	28	0.937	28	0.349
HRV	0.612	26	0.646	26	0.798	26	0.976	21	1	1	0.388
CZE	0.744	19	0.772	18	0.893	18	1	1	1	1	0.256
DNK	0.962	3	0.979	3	1	1	1	1	1	1	0.038
EST	0.866	8	0.887	9	0.971	8	1	1	1	1	0.134
FIN	0.955	4	0.968	4	1	1	1	1	1	1	0.045
FRA	0.859	10	0.885	10	0.966	9	1	1	1	1	0.141
DEU	0.798	14	0.827	14	0.936	14	1	1	1	1	0.202
GRC	0.572	28	0.605	28	0.752	28	0.925	27	0.962	26	0.390
HUN	0.691	21	0.725	21	0.843	21	0.961	23	0.976	24	0.285
ISL	0.901	7	0.922	7	0.993	7	1	1	1	1	0.099
IRL	0.792	15	0.821	15	0.940	13	1	1	1	1	0.208
ITA	0.619	25	0.659	25	0.798	25	0.937	25	0.962	25	0.343
LVA	0.700	20	0.733	20	0.868	20	0.996	18	1	1	0.300
LTU	0.749	17	0.788	17	0.951	12	1	1	1	1	0.251
LUX	0.850	11	0.874	11	0.957	11	1	1	1	1	0.150
NLD	0.909	6	0.930	6	1	1	1	1	1	1	0.091
NOR	0.913	5	0.933	5	0.995	6	1	1	1	1	0.087
POL	0.747	18	0.772	19	0.886	19	1	1	1	1	0.253
PRT	0.691	22	0.723	22	0.832	23	0.927	26	0.950	27	0.259
ROU	0.526	29	0.565	29	0.711	29	0.864	29	0.893	29	0.367
SVK	0.625	24	0.662	24	0.819	24	0.968	22	0.984	22	0.359
SVN	0.821	12	0.842	13	0.920	17	0.998	17	1	1	0.179
ESP	0.684	23	0.721	23	0.841	22	0.959	24	0.983	23	0.299
SWE	0.968	2	0.982	2	1	1	1	1	1	1	0.032
CHE	1	1	1	1	1	1	1	1	1	1	0
GBR	0.860	9	0.887	8	0.964	10	1	1	1	1	0.140

restrictive cases are defined as $\Delta CI = CI_1^D - CI_0^D$.

Table 4 shows that the unrestricted DEA model (full weight flexibility, $\sigma = 0$ and CI_0^D) has limited discrimination power. More than two-thirds of the countries (20 from 29 countries) obtain a CI score equal to 1, which is an evidence of the weak capacity to discriminate the performance of countries. The average of CI_0^D is also very close to 1 (0.988). The discrimination power progressively increases with the increase of σ, with only 16 countries obtaining a $CI_{0.1}^D$ score equal to 1 for $\sigma = 0.10$ (average 0.979), 5 countries obtaining a $CI_{0.5}^D$ equal to 1 and average 0.905 for $\sigma = 0.50$, and only 1 country obtaining a $CI_{0.9}^D$ score equal to 1 for $\sigma = 0.90$ (average 0.806) or $\sigma = 1$ (average 0.779). This highlights the critical role of weight restrictions to increase discrimination in performance assessments, particularly when the construction of rankings is important for public dissemination of the results.

Furthermore, countries with good performance only on a small subset of indicators will have CI sores that are more sensitive to the imposition of weight restrictions. Therefore, a comparative analysis of the results obtained for progressively stricter weight restrictions also identifies the countries with a more balanced profile regarding the multidimensional aspects of educational achievements. For example, the last column of Table 4, reporting the difference ΔCI between the CI_0^D score for $\sigma = 0$ (free weights) and CI_1^D score for $\sigma = 1$ (equal weights), shows that Switzerland is the country with the most stable CI score, meaning that it can be considered a benchmark to other countries irrespectively

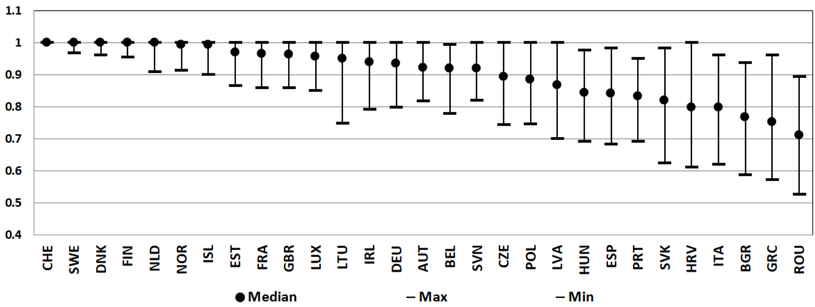

Figure 1: Results of the robustness analysis.

of the restrictions imposed on weights. Other countries, such as the Czech Republic or Lithuania, have a CI score equal to 1 in a scenario of free weights, but lose their benchmark status when the weight flexibility is reduced. This means that these countries may need to go an extra mile to achieve well balanced performance regarding all ET2020 strategic objectives.

Next, we explore the robustness of the countries CI scores in relation to variations in the value of $\sigma = 1$. Figure 1 shows the results of the countries' performance for different values of σ ranging between 0 and 1. Whiskers plot the median (circle), the maximum and minimum value of the CI scores distribution for each country. The countries are ordered in the x-axis by the median values.

Switzerland, Finland, Sweden and Denmark are unarguably the countries with the best education achievements. There are other countries, such as Slovakia, Italy, Croatia, Bulgaria, Greece and

Romania, whose relative performance is strongly influenced by the degree of importance attributed to each sub-indicator. For these six countries, represented on the right-hand side of Figure 2, the range between the maximum and minimum CI scores obtained for different values of σ is greater than 0.34. This suggests that their efforts to converge towards the EU education and training standards will be particularly demanding.

3.5 Use of the CI model to manage performance

For illustrative purposes, this section outlines the information that can be used to guide performance improvements. First, we identify the peers and target that are obtained using the CI model with different values of σ. We also present the potential for improvement at the system level and compare ET2020 targets with the efficient targets estimated using optimization.

Concerning the peers and targets, we use the case of Portugal as an illustrative example. Table 5 displays detailed information about the peers that are obtained for Portugal for different values of σ.

Using expression 5 and the values of λ_j for the peers reported in Table 5, it is possible to specify targets for improvement for inefficient countries, corresponding to a linear combination of sub-indicator values observed in peers, using the λ values as the coefficients of the aggregation. Table 6 illustrates the targets obtained for the assessment of Portugal for $\sigma = 0$. For the case of

Table 5: Peers underlying the performance evaluation of Portugal for different values of σ.

λ values for the peers identified						
Degree of flexibility	DNK	IRL	GBR	EST	NLD	CHE
$\sigma = 0$	0.2502	0.4890	0.2608	-	-	-
$\sigma = 0.1$	0.7919	-	-	0.0002	0.2079	-
$\sigma = 0.5$	0.7842	-	-	-	0.2065	0.0093
$\sigma = 0.9$	-	-	-	-	-	1
$\sigma = 1$	-	-	-	-	-	1

Table 6: Targets for Portugal, estimated with $\sigma = 0$ ($CI_0^D = 0.9503$).

	PRT		DNK	IRL	GBR
	Obs	Target (R + NR)	$\lambda = 0.2502$	$\lambda = 0.4890$	$\lambda = 0.2608$
Y_1	86.3	91.9 (90.8 + 1.1)	92.2	93.2	89.2
Y_2	31.9	50.7 (33.6 + 17.1)	47.6	53.8	47.9
Y_3	93.6	98.5 (98.5 + 0)	98.5	97.7	100
Y_4	72.2	80.9 (76.0 + 4.9)	81.7	77.9	85.7
Y_5	9.7	15.1 (10.2 + 4.9)	31.3	6.5	15.7
Y_6	80.5	84.7 (84.7 + 0)	85.2	86.5	80.9
Y_7	8.8	9.3 (9.3 + 0)	8.4	9.2	10.2
Y_8	45.1	80.6 (47.5 + 33.1)	84.0	81.1	79.7

Obs – observed values, R – radial adjustment, NR – non-radial adjustment.

the model with $\sigma = 0$, the targets are also decomposed in radial (R) and non-radial (NR) adjustments.

For this example, we can see that Y_3, Y_6 and Y_7 can improve only by 5.2%, whereas for the other indicators the potential for improvement is larger. In the case of indicators Y_2 ("the percentage of people aged 30-34 who have successfully completed tertiary education") and Y_8 ("the percentage of people aged 25-64 who have successfully completed at least upper secondary education"), the potential for non-radial improvement is particularly significant, meaning that these are the aspects in which Portugal shows worse performance in comparison with its peers. Table 7 illustrates the targets obtained for the assessment of Portugal for different levels of weight flexibility.

For the model with complete flexibility ($\sigma = 0$), the targets dominate the observed achievement in all dimensions (sub-indicators), which does not occur for the fixed weight case ($\sigma = 1$), when the weights are fixed, only one country can be used as peer, such that if it has poor performance in one dimension (such as Y_3 in the case of Switzerland), the targets for inefficient countries may even suggest a decline in performance in this dimension (as shown in this example, with a target for Y_3 corresponding 81.3, when the observed level in Portugal is 93.6). This shows that the DEA model with fixed weights (or equivalently the SAW model) may be inadequate for target setting purposes.

In conclusion, a drawback of the fixed weighting system is that it relies too much on a single observation. Conversely, an evaluation

Table 7: Targets for Portugal, estimated for different values of σ.

		$\sigma = 0.1$		$\sigma = 0.5$		$\sigma = 0.9$ ir $\sigma = 1$	
	Obs	T	R+NR	T	R+NR	T	R+NR
Y_1	86.3	92.1	90.8+1.3	92.1	90.8+1.3	94.8	90.8+4
Y_2	31.9	47.3	33.6+13.8	47.4	33.6+13.8	49.3	33.6+15.7
Y_3	93.6	98.3	98.5-0.2	98.2	98.5-0.3	81.3	98.5-17.2
Y_4	72.2	83.1	76+7.1	83.1	76+7.1	84.6	76+8.6
Y_5	9.7	28.7	10.2+18.5	28.7	10.2+18.5	30.8	10.2+20.6
Y_6	80.5	84.6	84.7-0,1	84.6	84.7-0.2	81.9	84.7-2.8
Y_7	8.8	9.3	9.3+0	9.3	9.3+0	12.3	9.3+3
Y_8	45.1	79.6	47.5+32.1	79.6	47.5+32.2	87.3	47.5+39.8

Obs – observed values, T – target, R – radial adjustment, NR – non-radial adjustment.

with some degree of flexibility can be a balanced compromise, enabling more insightful identification of peers and targets. The imposition of weight restrictions reduces the set of peers that can be used as benchmarks for all other countries, but each inefficient country can choose a linear combination of some of these countries to define its own targets. In this way, the targets are more aligned with the countries' own priorities than in the case of using a single benchmark, identified using the fixed weights approach.

Regarding the conclusions of this study for policy purposes, we highlight the following: first, there are some countries (Switzerland, Sweden, Denmark and Finland) that seem to be highly efficient regardless of the degree of flexibility allowed with respect to the importance of each sub-indicator. This core group is aligned with the previous finding of the education efficiency measurement literature. Finland was considered a benchmark by Clements [23], Sutherland et al. [53], Giambona et al. [33], Thieme et al. [55]; Agasisti [6] and Bogetoft et al. [15]; Switzerland was found to be a benchmark by Agasisti [5, 6], and Sweden by Afonso and St. Aubyn [4]. There are other countries (Slovakia, Italy, Croatia, Bulgaria, Greece and Romania) whose performance is strongly influenced by the degree of flexibility allowed to the weights, signalling uneven performance in different dimensions of education achievements. This suggests that their efforts to converge towards the EU education and training standards will be particularly demanding.

Countries with good performance only on a small subset of indicators are more sensitive to the degree of weight restrictions flexibility. For example, Switzerland is the country with the most stable efficiency score, so it can be considered a benchmark to other countries irrespectively of the restrictions imposed on weights. Other countries (Czech Republic and Lithuania) are efficient in a scenario of free weights, but lose their benchmark status when the weight flexibility is reduced, meaning that they may need to go an extra mile to achieve well balanced performance.

4 Analysis of data by means of PCA-DEA

The other way dealing with discrimination issue the dimension reduction of sub-indicators before running DEA. Principal components analysis (PCA) could be employed where the original number of indicators would be replaced by a smaller number of principal components with a minimal loss of information [1]. We have applied the algorithm of PCA to the data. Since the first four components explain about 89.42% of variance (PC1 – 46.16%, PC2 – 24.34%, PC3 – 11.66% and PC4 – 7.26%) they are considered as representing the data sufficiently well.

Further we apply DEA for the projection of the original data to the subspace defined by four eigenvectors computed by means of PCA. For details we refer to [1] where an excellent description of the algorithm is presented. The attention should be paid that some elements of reduced data sets are negative which seem not natural in the context of the original formulation of the DEA problems. However, alternative DEA models have been developed which are translation invariant [7, 13]. We will use the so-called Additive Model which maintains translation invariance in the case of the output data of analysis alone [7]. Since we consider the output data only, the Additive Model is defined as the following problem of linear programming:

$$\max z_{jO} = \sum_{k=1}^{d} s_k^+, \tag{11}$$
$$\sum_{j=1}^{n} \lambda_j \widetilde{p}_{kj} = \widetilde{p}_{kjO} + s_k^+,$$
$$\sum_{j=1}^{n} \lambda_j = 1,$$
$$s_k^+ \geq 0.$$

where $p_{kj}, k = 1, \ldots, d, j = 1, \cdots, n$ – principal components, $\widetilde{p}_{kj} = p_{kj} + q$ and $q = -\min_{k,j} p_{kj} + 1$.

The solution of (11) z_i is equal to the L_1 (city block) distance from the DMU to the efficiency frontier. Thus, for the efficient DMUs $z_i = 0$. Correspondingly, the non-zero distance is a measure of inefficiency. The values of z_i for all DMUs are summarized in Table 8.

The distance from DMU to the efficiency frontier can be used as a criterion for ranking the inefficient DMUs. The non-zero slacks show the potential improvement quantities along directions of principal components. However, these potential efficiency improvements do not have proper interpretation. To obtain estimates of potential efficiency improvements with respect to original criteria the slacks can be expressed in terms of original data using loadings of principal components. However, this problem has

Table 8: The distances to the efficiency frontier.

Cnt	z_i	Cnt	z_i	Cnt	z_i	Cnt	z_i
PRT	0	CHE	0	LTU	13.1	DEU	29.9
ESP	0	SWE	0	FRA	13.7	ITA	34.3
GRC	0	FIN	0	LUX	14	HUN	45.1
ISL	0	EST	0	BEL	15.6	CZE	46.5
NOR	0	NLD	2.4	POL	19.4	HRV	53.9
IRL	0	GBR	4.6	LVA	25.7	SVK	69.2
DNK	0	SVN	11.6	AUT	28.7	BGR	72.4
						ROU	79.8

no ambiguous solution. For example, the PCA coordinates of vectors in the original space are obtained by linear projecting where a "shadow effect" can emerge. The mutual distances of images can be better preserved using non-linear projecting methods, e.g. Multidimensional Scaling (MDS) [64]. The attractive idea of hybridization of DEA and MDS, however, is very new and still not mature.

5 Methodology for effectiveness estimation of education systems

The methodology for effectiveness estimation of education systems was proposed according CI paradigm. The construction of a CI consists of several stages. The latter includes the selection of

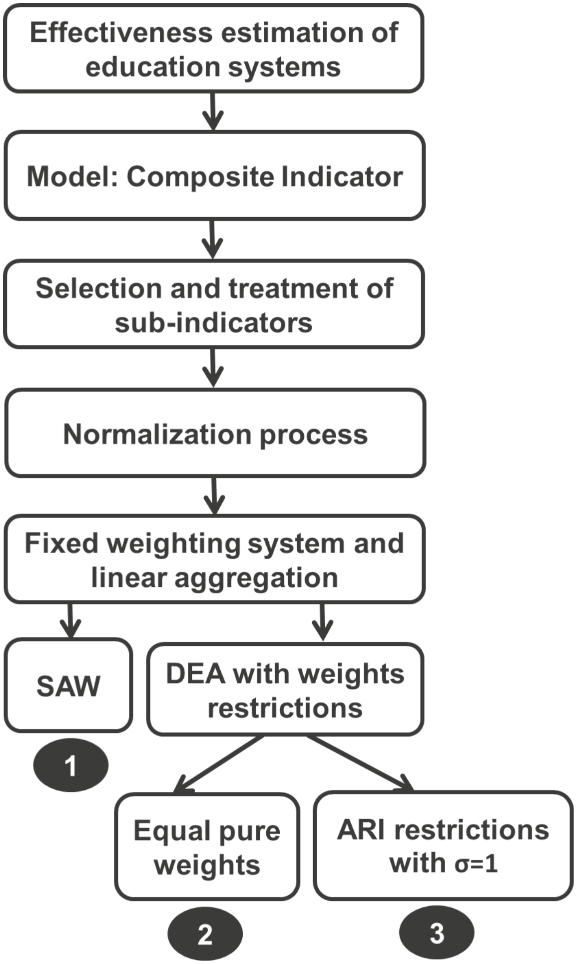

Figure 2: Methodology for effectiveness estimation of education systems with purpose to rank the countries.

sub-indicators and the treatment of data, the application of a normalization process (if needed), the specification of the weights for the sub-indicators and the selection of the aggregation function. The assignment of the weights and aggregation depend on the purpose of the CI.

When the CI is constructed for the purpose of effectiveness assessment (control) and countries' ranking, one can use the fixed weighting system (countries have the same set of weights). Therefore one can develop the CI implementing (Figure 2): (1) SAW model with equal weights, (2) DEA model with equal pure weights, (3) DEA model with ARI weight restrictions and parameter $\sigma = 1$. In all cases, linear aggregation is applied. The calculated CI, irrespectively on the method chosen, will give identical countries' ranking.

When the CI is constructed for the purpose is effectiveness management (improvement), one can use the flexible weighting system (set of weights is different across DMUs). For the development of the CI one can implement (Figure 3) DEA model with ARI weight restrictions and parameter $\sigma \in [0; 1)$. The selection value of the parameter σ will depend on the data – as number of the sub-indicators are similar to the number of DMU, the value of σ should be larger to ensure sufficient discrimination power of the DEA model.

If sub-indicators are disproportionately compared to the number of DMUs, it is possible to reduce the data dimension before running the DEA model. This can be implemented (Figure 3)

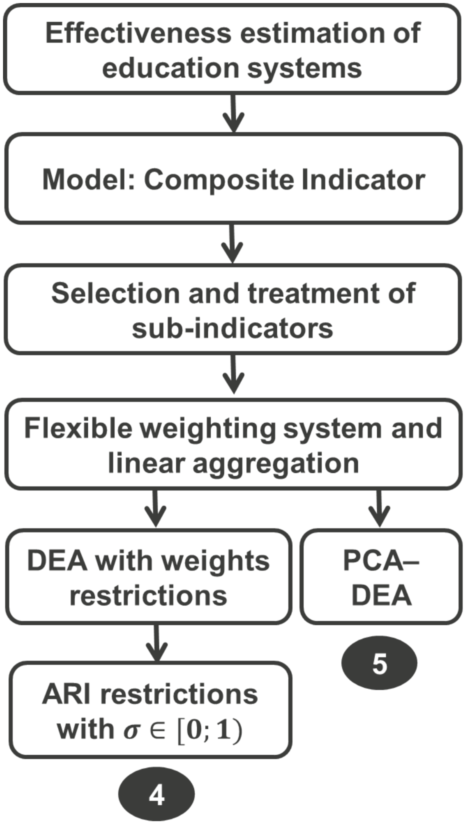

Figure 3: Methodology for effectiveness estimation of education systems with purpose to improve the countries' education systems.

applying (5) hybrid PCA-DEA model.

6 The main resuls and conclusions

The CIs have been accepted as a useful tool for conducting performance assessment and construction of rankings in various fields, but only a few CIs were constructed in the area of education and the Lithuanian education system was not evaluated in this context. The literature review on frontier-based efficiency studies represents that only a small number of efficiency studies of the education sector focused at country level (DMU is a country) assessments. Furthermore, most studies were focused exclusively on primary, secondary or tertiary education stages, without providing an encompassing picture of the entire education system.

The ratio of sub-indicators (inputs $m = 1$ and outputs $s = 8$) and DMUs ($n = 29$) is satisfied ($n \geq 3\dot{(}m \cdot s)$ and $n \geq 3\dot{(}m + s)$) in the construction of CIs for the evaluation of European countries' education systems, but the discrimination power of the DEA model is insufficient to compute the CIs and rank the countries' education systems. According to experimental research, the maximum number of outputs can be $s = 3$ or the ratio of sub-indicators and DMUs is maintained ($n \geq 6\dot{(}m + s)$ or $n \geq 8\dot{(}m \cdot s)$); only then provides the DEA model with sufficient the discrimination power to rank the European countries' education systems.

The unrestricted DEA model showed limited discrimination power in this study – more than two-thirds of the countries are Pareto optimal (in 2013 – 18 countries, 2014 and 2015 – 20 countries out of the 29 countries) and their CIs are equal to 1. Moreover using the unrestricted DEA model to construct CI, a large number of the weights are equal to 0, so these sub-indicators are ignored in the estimation of the CI score (sub-indicators to be assigned a zero weight are not included in the CI). An important procedure to increase the discrimination power of the model and to avoid the occurrence of weights equal to zero is the use of weight restrictions.

The SAW model with equal weights imposed on normalised data gives exactly the same ranking as the DEA model with equal pure weights imposed on normalised data ($R^S_{abs} = R^D_{abs}$). This confirms the possibility of using DEA models for constructing CIs as an alternative to the traditional SAW model. In the case of the DEA model the weights have not been decided a-priori and it is possible to specify weights for each sub-indicator that are different across DMU, can be used data measured in different scales (avoiding the need to normalise data), the CI score provides a relative measure of performance that facilities the interpretation of the results of the assessment; peers and targets are also sought, contributing to a more informed design of policies to improve performance.

Further the Spearman correlation coefficient between the ranks of the DEA model with equal pure weights and the DEA model with ARI restrictions is 0.942, the difference of medians is 0 (the Wilcoxon's signed-rank test for matched pairs the null hypothesis is not rejected, p-value = 0.905) and the similarity between the average is remarkable (0.771 versus 0.779). This similarity in the results of these two modelling alternatives opens up the possibility of using DEA models with ARI weight restrictions to progressively depart from fixed weighting to flexible weighting systems.

The main contribution of this study is the development of a new type of weight restriction that gradually allows departing from fixed to flexible weighting systems, enhances discrimination power and avoids the occurrence of zero weights in the DEA framework. The weight flexibility was implemented through the new parameter $\sigma \in [0;1]$. When the parameter $\sigma = 1$, the DEA model corresponds to a formulation with equal weights for all dimensions (representing a fixed weights scenario). When the parameter $\sigma = 0$, the least restrictive case corresponds to the DEA model without weight restrictions. Intermediate scenarios can be obtained using other values of $\sigma \in (0;1)$.

The DEA model with fixed weights may be inadequate for target setting purposes. The main drawback of the fixed weighting system is that it relies too much in a single observation, so an evaluation with some degree of flexibility can be a balanced compromise, enabling a more insightful identification of peers and targets. The imposition of flexible weight restrictions reduces the set of peers that can be used as benchmarks, but each inefficient country can choose a linear combination of some of these countries to define its own targets. In this way, the targets are more aligned with the countries' own priorities than in the case of using a single benchmark identified using the fixed weights approach. As is shown in this study the DEA model can be a useful tool for fine tuning the EU strategy, based on the current level of achievements observed in the EU countries.

The use of PCA-DEA approach for performance assessment has been investigated in this study (only two studies in academic literature have used this approach for education data). The discrimination power of the DEA model has increased – instead of two-thirds of efficient countries, one-third of the countries' education systems has been as Pareto optimal (8 instead of 18 countries in 2013, 10 instead of 20 countries in 2014 and 2015), but this is not enough for ranking the EU countries. According to all this research, the methodology for performance evaluation of education systems was proposed.

The following conclusions can be drawn from this study:

1. The ratio of sub-indicators and DMUs is satisfied in the construction of CIs for the evaluation of European countries' education systems, but the discrimination power of the DEA model is insufficient to compute the CIs and rank the countries' education systems. Moreover a large number of the weights are equal to 0, so these sub-indicators are ignored in the estimation of the CI score. For obtaining a higher discrimination power and non-zero weights, weight restrictions should be applied in the DEA model.

2. The SAW model with equal weights gives exactly the same

ranking as the DEA model with equal weights. This confirms the possibility of using DEA models for constructing CIs as an alternative to the traditional SAW model.

3. The similarity between the DEA model with equal pure weights and the DEA model with ARI restrictions opens up the possibility of using DEA models with ARI weight restrictions to transition from fixed to flexible weighting systems.

4. The DEA model with ARI weight restrictions and the new flexibility parameter $\sigma \in [0; 1]$ allows gradual transition from fixed to flexible weighting systems. The model solves the problem of insufficient discrimination of the DEA model, insures the inclusion of all sub-indicators in the CI and provides additional information that may be used to improve system performance.

5. Despite the fact that the hybrid PCA-DEA model has a higher discrimination power than the unrestricted DEA model when assessing the European countries' education systems, it has remained insufficient to perform ranking of the countries analysed.

List of Publications on the Topic of Dissertation

1. *Stumbrienė, D., Camanho, A. S., Jakaitienė, A. (). The performance of education systems in the light of Europe 2020 strategy. Annals of Operations Research,* (under review) [WoS]

2. Jakaitienė, A., Žilinskas, A., Stumbrienė, D. (2018). Analysis of Education Systems Performance in European Countries by Means of PCA-DEA. Informatics in Education, 17(2), 245-263. [Emerging WoS]

3. Stumbrienė, D., Jakaitienė, A., Želvys, R. (2017). Švietimo sistemos stebėsena: išteklių ir rezultatų indeksų sąveika. Lietuvos statistikos darbai, 56(1).

4. Želvys R., Jakaitienė A., Stumbrienė D. (2017). Moving towards different educational models of the welfare state: comparing the education systems of the Baltic countries. Filosofija. Sociologija. ISSN 0235-7186. Nr.2. [WoS]

References

[1] Nicole Adler and Boaz Golany. Evaluation of Deregulated Airline Networks Using Data Envelopment Analysis Combined with Principal Component Analysis with an Application to Western Europe. *European Journal of Operational Research*, 132(2):260–273, 2001.

[2] Nicole Adler and Boaz Golany. Including Principal Component Weights to Improve Discrimination in Data Envelopment Analysis. *Journal of the Operational Research Society*, 53(9): 985–991, 2002.

[3] António Afonso and Miguel St Aubyn. Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic modelling*, 23(3):476–491, 2006.

[4] António Afonso and Miguel St Aubyn. Non-parametric approaches to education and health expenditure efficiency in oecd countries. 2004.

[5] Tommaso Agasisti. Performances and spending efficiency in higher education: a european comparison through non-parametric approaches. *Education Economics*, 19(2):199–224, 2011.

[6] Tommaso Agasisti. The efficiency of public spending on education: an empirical comparison of eu countries. *European Journal of Education*, 49(4):543–557, 2014.

[7] Agha Iqbal Ali and Lawrence M Seiford. Translation Invariance in Data Envelopment Analysis. *Operations Research Letters*, 9(6):403–405, 1990.

[8] Robert Allen, Antreas Athanassopoulos, Robert G Dyson, and Emmanuel Thanassoulis. Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. *Annals of operations research*, 73:13–34, 1997.

[9] Milan Andrejić, Nebojša Bojović, and Milorad Kilibarda. Benchmarking Distribution Centres Using Principal Component Analysis and Data Envelopment Analysis: a Case Study of Serbia. *Expert Systems with Applications*, 40(10): 3926–3933, 2013.

[10] D Annapoorni and V Prakash. Measuring the Performance Efficiency of Hospitals: PCA-DEA Combined Model Approach. *Indian Journal of Science and Technology*, 9(S1), 2017.

[11] A Azadeh, SF Ghaderi, and V Ebrahimipour. An Integrated PCA DEA Framework for Assessment and Ranking of Manufacturing Systems Based on Equipment Performance. *Engineering Computations*, 24(4):347–372, 2007.

[12] Rajiv D Banker and Srikant M Datar. Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research*, pages 21–39, 1989.

[13] Rajiv D Banker, Abraham Charnes, and William Wager Cooper. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092, 1984.

[14] Miroslav Beblavy, Anna-Elisabeth Thum, and Marcela Veselkova. Education Policy and Welfare Regimes in OECD Countries: Social Stratification and Equal Opportunity in Education. 2011.

[15] Peter Bogetoft, Eskil Heinesen, and Torben Tranæs. The efficiency of educational production: A comparison of the nordic countries with other oecd countries. *Economic Modelling*, 50: 310–321, 2015.

[16] William F Bowlin. Measuring performance: An introduction to data envelopment analysis (dea). *The Journal of Cost Analysis*, 15(2):3–27, 1998.

[17] Josip Burušić, Toni Babarović, and Marija Šakić Velić. School Effectiveness: an Overview of Conceptual, Methodological and Empirical Foundations. In *School Effectiveness and Educational Management*, pages 5–26. Springer, 2016.

[18] Felipe A Calabria, Ana S Camanho, and Andreia Zanella. The use of composite indicators to evaluate the performance of brazilian hydropower plants. *International Transactions in Operational Research*, 25(4):1323–1343, 2018.

[19] Abraham Charnes, William Wager Cooper, and Edwardo

Rhodes. Measuring the Efficiency of Decision-Making Units. *European Journal of Operational Research*, 3(4):339, 1979.

[20] Jihong Chen, Zheng Wan, Fangwei Zhang, Nam-kyu Park, Xinhua He, and Weiyong Yin. Operational Efficiency Evaluation of Iron Ore Logistics at the Ports of Bohai Bay in China: Based on the PCA-DEA Model. *Mathematical Problems in Engineering*, 2016, 2016.

[21] Laurens Cherchye, Willem Moesen, Nicky Rogge, and Tom Van Puyenbroeck. An introduction to 'benefit of the doubt'composite indicators. *Social indicators research*, 82(1): 111–145, 2007.

[22] Laurens Cherchye, Wim Moesen, Nicky Rogge, Tom Van Puyenbroeck, Michaela Saisana, Andrea Saltelli, Roman Liska, and Stefano Tarantola. Creating composite indicators with dea and robustness analysis: the case of the technology achievement index. *Journal of the Operational Research Society*, 59(2):239–251, 2008.

[23] Benedict Clements. How efficient is education spending in europe? *European Review of Economics and Finance*, 1(1): 3–26, 2002.

[24] Joint Research Centre-European Commission et al. *Handbook on constructing composite indicators: Methodology and user guide*. OECD publishing, 2008.

[25] William W Cooper, Lawrence M Seiford, and Kaoru Tone. *Introduction to Data Envelopment Analysis and its Uses:*

with DEA-Solver Software and References. Springer Science & Business Media, 2006.

[26] JM Cordero-Ferrera, Francisco Pedraja-Chaparro, and D Santín-González. Enhancing the inclusion of non-discretionary inputs in dea. *Journal of the Operational Research Society*, 61(4):574–584, 2010.

[27] Fengxia Dong, Paul D Mitchell, Deana Knuteson, Jeffery Wyman, AJ Bussan, and Shawn Conley. Assessing Sustainability and Improvements in US Midwestern Soy Bean Production Systems Using a PCA-DEA Approach. *Renewable Agriculture and Food Systems*, 31(6):524–539, 2016.

[28] Robert G Dyson and Emmanuel Thanassoulis. Reducing weight flexibility in data envelopment analysis. *Journal of the operational research society*, 39(6):563–576, 1988.

[29] Robert G Dyson, Emmanuel Thanassoulis, and Aziz Boussofiane. Data envelopment analysis. *Operational research tutorial papers*, pages 13–28, 1990.

[30] Robert G Dyson, Rachel Allen, Ana S Camanho, Victor V Podinovski, Claudia S Sarrico, and Estelle A Shale. Pitfalls and Protocols in DEA. *European Journal of Operational Research*, 132(2):245–259, 2001.

[31] S Flisi, V Goglio, and E Meroni. Monitoring the Evolution of Education and Training Systems: a Guide to the Joint Assessment Framework. *Joint Research Center. Publications Office of the European Union. Luxembourg*, 10:15187, 2014.

[32] Harold O Fried, CA Knox Lovell, and Shelton S Schmidt. *The Measurement of Productive efficiency and Productivity Growth*. Oxford University Press, 2008.

[33] Francesca Giambona, Erasmo Vassallo, and Elli Vassiliadis. Educational systems efficiency in european union countries. *Studies in Educational Evaluation*, 37(2-3):108–122, 2011.

[34] Víctor Giménez, Diego Prior, and Claudio Thieme. Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. *Journal of the Operational Research Society*, 58(8):996–1007, 2007.

[35] Hariolf Grupp and Mary Ellen Mogee. Indicators for national science and technology policy: how robust are composite indicators? *Research Policy*, 33(9):1373–1384, 2004.

[36] Eric A Hanushek and Javier A Luque. Efficiency and equity in schools around the world. *Economics of education Review*, 22(5):481–502, 2003.

[37] Isabel M Horta, Ana S Camanho, and Jorge Moreira Da Costa. Performance assessment of construction companies integrating key performance indicators and data envelopment analysis. *Journal of Construction Engineering and Management*, 136(5):581–594, 2009.

[38] Changbing Jiang and Peihua Fu. Evaluating Efficiency and Effectiveness of Logistics Infrastructure Based on PCA-DEA Approach in China. In *Intelligent Computation Technology*

and Automation, 2009. ICICTA'09. Second International Conference, volume 3, pages 62–66. IEEE, 2009.

[39] Dhanya Jothimani, Ravi Shankar, and Surendra S Yadav. A PCA-DEA Framework for Stock Selection in Indian Stock Market. *Journal of Modelling in Management*, 12(3):386–403, 2017.

[40] Liang Liang, Yongjun Li, and Shibing Li. Increasing the Discriminatory Power of DEA in the Presence of the Undesirable Outputs and Large Dimensionality of Data Sets with PCA. *Expert Systems with Applications*, 36(3):5895–5899, 2009.

[41] Steve Wenbin Liu, Wei Meng, and TQ Zhang. Incorporating Value Judgments in DEA. 2006.

[42] Paulo Morais and Ana S Camanho. Evaluation of performance of european cities with the aim to promote quality of life improvements. *Omega*, 39(4):398–409, 2011.

[43] Pilar Murias, José Carlos de Miguel, and David Rodríguez. A composite indicator for university quality assesment: The case of spanish higher education system. *Social Indicators Research*, 89(1):129–146, 2008.

[44] Maria Nieswand, Astrid Cullmann, and Anne Neumann. Overcoming Data Limitations in Non-parametric Benchmarking: Applying PCA-DEA to Natural Gas Transmission. 2009.

[45] VV Podinovski and AD Athanassopoulos. Assessing the relative efficiency of decision making units using dea models with weight restrictions. *Journal of the Operational Research Society*, 49(5):500–508, 1998.

[46] Reet Põldaru, Jüri Roots, et al. Assessing the Performance of Milk Producer in Estonian Farms Using PCA-DEA Approach. *Production and Co-operation in Agriculture*, 2014.

[47] Andrea Saltelli. Composite indicators between analysis and advocacy. *Social indicators research*, 81(1):65–77, 2007.

[48] Cláudia S Sarrico and RG Dyson. Restricting virtual weights in data envelopment analysis. *European Journal of Operational Research*, 159(1):17–34, 2004.

[49] M.C.A Silva and A.S. Camanho. Using Data Analytics to Benchmark Schools: the Case of Portugal. In Jan Vanthienen and Kristof De Witte, editors, *Analytics Applications in Education*, chapter 6, pages 129–192. Taylor and Francis, 2017.

[50] Zilla Sinuany-Stern and Lea Friedman. Dea and the discriminant analysis of ratios for ranking units. *European Journal of Operational Research*, 111(3):470–478, 1998.

[51] Amina Ahec Šonje, Milan Deskar-Škrbić, and Velimir Šonje. Efficiency of public expenditure on education: Comparing croatia with other nms. In *12th Annual International Technology, Education and Development Conference*, 2018.

[52] Bertelsmann Stiftung. *The German Learning Atlas: Making lifelong learning tangible on a regional level.* Bertelsmann Stiftung, 2013.

[53] Douglas Sutherland, Robert Price, and Frédéric Gonand. Improving public spending efficiency in primary and secondary education. *OECD Journal: Economic Studies*, 2009(1):1–30, 2010.

[54] Emmanuel Thanassoulis, Maria Conceição Portela, and Rachel Allen. Incorporating value judgments in dea. In *Handbook on data envelopment analysis*, pages 99–138. Springer, 2004.

[55] Claudio Thieme, Víctor Giménez, and Diego Prior. A comparative analysis of the efficiency of national education systems. *Asia Pacific Education Review*, 13(1):1–15, 2012.

[56] Russell G Thompson, FD Singleton Jr, Robert M Thrall, and Barton A Smith. Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas. *Interfaces*, 16(6):35–49, 1986.

[57] Russell G Thompson, Larry N Langemeier, Chih-Tah Lee, Euntaik Lee, and Robert M Thrall. The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming. *Journal of Econometrics*, 46(1-2):93–108, 1990.

[58] Tohru Ueda and Yoko Hoshiai. Application of Principal Component Analysis for Parsimonious Summarization of

DEA Inputs and/or Outputs. *Journal of the Operations Research Society of Japan*, 40(4):466–478, 1997.

[59] Kristof De Witte and Laura López-Torres. Efficiency in Education: a Review of Literature and a Way Forward. *Journal of the Operational Research Society*, 68(4):339–363, 2017.

[60] Y-HB Wong and JE Beasley. Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 41(9):829–835, 1990.

[61] Andreia Zanella, AS Camanho, and Teresa Galvão Dias. Benchmarking countries' environmental performance. *Journal of the Operational Research Society*, 64(3):426–438, 2013.

[62] Andreia Zanella, Ana S Camanho, and Teresa G Dias. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research*, 245(2):517–530, 2015.

[63] Rimantas Želvys, Audronė Jakaitienė, and Dovilė Stumbrienė. Moving Towards Different Educational Models of the Welfare State: Comparing the Education Systems of the Baltic Countries. *Filosofija. Sociologija*, 28(2):139–150, 2017.

[64] Antanas Žilinskas and Julius Žilinskas. Parallel hybrid algorithm for global optimization of problems occurring in mds-based visualization. *Computers & Mathematics with Applications*, 52(1-2):211–224, 2006.

Duomenų apgaubties metodas švietimo sistemų efektyvumo analizėje

Santrauka

Tyrimo objektas

Švietimas yra daugialypė kompleksinė sistema, kurią sudaro įvairaus dydžio ir sudėtingumo posistemiai (ikimokyklinis ugdymas, bendrasis ugdymas, profesinis rengimas, aukštasis mokslas, suaugusiųjų mokymasis ir kt.), o švietimo būklę nusako ir rezultatus lemia daug skirtingų veiksnių. Sistemos kompleksiškumas kuria erdvę įvairaus tipo uždaviniams, ypač kai reikia atlikti švietimo sistemos vertinimą ir skirtingų šalių sistemų palyginimą. Įvairių šalių švietimo sistemų palyginimas yra sudėtingas daugiakriterinis optimizavimo uždavinys.

Kai analizuojama kompleksinė sistema ir pavienių rodiklių analizė yra ribota, įprasta skaičiuoti sudėtinius rodiklius, siekiant apibendrinus individualius rodiklius gauti vieną agreguotą rodiklį, nusakantį visos sistemos būklę. Paprasčiausias būdas agreguoti rodiklius – paprastasis adityvusis svorinis metodas (SAW, angl. *Simple Additive Weighting*), kai naudojami vienodi svoriai. Tačiau dėl skirtingos aplinkos ir sąlygų Europos šalių švietimo sistemų būklė ir pasiekimai yra nevienodi, todėl vertinti jų remiantis vienodais kriterijais nėra tikslinga. Be to, kaip pažymėjo Silva ir kt. (2017) [49], sistemos efektyvumo vertinimas absoliučiais skaičiais dažniausiai nėra toks vertingas kaip palyginimas su kitomis sistemomis – gerosios praktikos pavyzdžių pateikimas gali tapti sistemų tobulinimo pagrindu. Šio darbo pagrindinis objektas – duomenų apgaubties analizė (DEA, angl. *Data Envelopment Analysis*) ir jos taikymas 29 Europos šalių švietimo sistemų efektyvumui vertinti, naudojant 2013–2015 m. Eurostato ir EBPO duomenis.

Darbo aktualumas

Dėl švietimo sistemos kompleksiškumo ir sudėtingumo mokslinėje literatūroje tik nedaugelyje tyrimų atliekama šalies lygmens analizė (sprendimų priėmimo vienetas – šalis), tačiau mokyklos lygmens analizė yra plačiai tiriama tema [59]. Gali būti išskirta keletas švietimo efektyvumo tyrimų ([3–6, 23, 34, 36, 51, 53, 55]), kuriuose yra atlikta šalies lygmens analizė, tačiau šiuose tyrimuose analizuojama ne visa šalies švietimo sistema, o jos posistemiai (pradinis, pagrindinis ar vidurinis ugdymas, aukštasis mokslas). Atskirų švietimo posistemių analizė neatspindi visos švietimo sistemos, tik atlikus visų švietimo posistemių analizę, galima vertinti šalies švietimo sistemos būklę. Turimomis žiniomis, tik Bogetoft ir kt. (2015) [15] analizavo švietimo sistemų kaip visumos (pradinio, pagrindinio ir vidurinio ugdymo bei aukštojo mokslo) efektyvumą ir palygino skirtingų šalių duomenis. Galima teigti, kad švietimo sistemos (kaip visumos) analizė yra mažai nagrinėta mokslinių tyrimų sritis.

Pastaraisiais metais rodikliams agreguoti vis dažniau naudojamas DEA metodas [18, 22, 37, 42, 43, 61, 62], kurį taikant atliekamas lanksčiųjų svorių priskyrimas rodikliams. Tačiau dėl lengvo pritaikomumo ir skaidrumo nepamirštamas ir SAW metodas, pagal kurį atliekamas fiksuotųjų svorių priskyrimas rodikliams. Vi dėlto mokslinėje literatūroje nėra aprašyto laipsniško perėjimo nuo fiksuotųjų svorių priskyrimo prie lanksčiųjų, t. y. šie metodai nėra susieti.

Lankstumas parenkant svorius yra vienas iš pagrindinių DEA metodo privalumų, tačiau, esant gana dideliam rodiklių skaičiui ir palyginti mažam vertinamų šalių skaičiui, kai kuriems rodikliams gali būti priskirti nuliniai svoriai ir DEA skiriamoji geba (angl. *discrimination power*) sumažėja, todėl didelė dalis šalių švietimo sistemų gali būti įvertintos kaip efektyviai veikiančios [41], o tai yra nepageidaujama, sprendžiant šalių rangavimo uždavinius.

Mokslinėje literatūroje [12, 16, 29, 30, 50] yra aprašytos vertinamų rodiklių ir sprendimų priėmimo vienetų (DMU, angl. *Decision Making Unit*) skaičiaus santykio proporcijos, kurios turėtų būti išlaikytos taikant DEA metodą, kai analizuojamas sunaudotų išteklių pavertimas rezultatais, tačiau mokslinėje literatūroje nėra aprašyta vertinamų rodiklių ir DMU skaičiaus santykio proporcijos, kai skaičiuojamas sudėtinis rodiklis ir atliekamas DMU rangavimas. Šio tipo uždaviniams, kai turimi tik rezultatų rodikliai, o išteklių rodikliai yra fiktyvūs, būtina maksimali modelio skiriamoji geba, kad būtų galima atlikti šalių rangavimą.

Siekiant atlikti efektyviai veikiančių sistemų šalių rangavimą, į DEA modelį gali būti įtraukiamos papildomos prielaidos. Mokslinėje literatūroje [28, 30, 48, 54, 56, 57, 60, 62] dažniausiai pasitaikanti DEA modelio modifikacija, didinanti skiriamąją gebą,

– svorių apribojimų įtraukimas. Turimomis žiniomis, kol kas nėra pasiūlyta dalinio svorių lankstumo išraiška, leidžianti palaipsniui pereiti nuo fiksuotųjų prie lanksčiųjų svorių priskyrimo DEA modelyje ir užtikrinanti pakankamą DEA modelio skiriamąją galią bei visų rodiklių įtraukimą į modelį.

Kai turimų rodiklių skaičius, lyginant su vertinamų DMU skaičiumi, stipriai išaugs, patraukli alternatyva, padedanti spręsti DEA modelio skiriamosios galios problemą, bus duomenų dimensijos mažinimas prieš atliekant DEA. Darbuose [1, 2, 58] pagrįsta idėja sujungti PCA ir DEA yra gana retai taikoma (rasta vos keliolika mokslinių darbų), atliekant empirinius tyrimus aviacijos [1], gamybos [11, 46], logistikos [9, 20, 38], ekologijos [40, 44], žemdirbystės [27], finansų [39] ir sveikatos apsaugos [10] srityse. Turimomis žiniomis, yra tik vienas darbas, kuriame PCA-DEA modelis taikomas švietimo duomenims, tai pradinis Adler ir Golany [2] tyrimas, kuriame buvo vertinami septynių universitetų padalinių veiklos rezultatai.

Darbo tikslas ir uždaviniai

Darbo tikslas – pasiūlyti švietimo sistemų efektyvumo vertinimo metodiką, paremtą duomenų apgaubties analize.

Tikslui pasiekti sprendžiami šie uždaviniai:

1. Atlikti švietimo sistemų efektyvumo vertinimo tyrimų analizę, išskirtinį dėmesį skiriant darbams, kuriuose atliekama šalies lygmens analizė.

2. Taikant sudėtinių rodiklių metodiką, agreguoti atrinktus švietimo sistemos vertinimo rodiklius.
3. Pritaikyti DEA modelį visos švietimo sistemos efektyvumui vertinti ir šalies lygmens analizei atlikti.
4. Pasiūlyti DEA modelio modifikaciją, kaip laipsniškai pereiti nuo fiksuotųjų prie lanksčiųjų svorių.
5. Įvertinti duomenų dimensijos mažinimo metodo tinkamumą pakankamai DEA modelio skiriamajai galiai užtikrinti.
6. Remiantis atliktais tyrimais, pasiūlyti švietimo sistemų vertinimo metodiką.

Mokslinis darbo naujumas

Sudėtinių rodiklių skaičiavimas yra dažnai naudojamas, siekiant kiekybiškai įvertinti socialinius ir ekonominius reiškinius, tačiau pasaulyje švietimo procesams įvertinti buvo skaičiuojami vos keli sudėtiniai rodikliai. Lietuvos švietimo sistema šiame kontekste nebuvo vertinta nei vieną kartą. Šiame darbe sukurtas naujas sudėtinis rodiklis Europos šalių švietimo sistemų efektyvumui vertinti, taikant DEA metodiką, analizuojant visą švietimo sistemą, o ne atskirus jos posistemius, ir atliekant šalies lygmens analizę. Skaičiuojant Europos šalių švietimo sistemų sudėtinius rodiklius, parodyta, kad mokslinėje literatūroje pateiktos rodiklių ir vertinamų šalių skaičiaus santykio proporcijos yra nepakankamos, kai skaičiuojamas sudėtinis rodiklis, skirtas šalių rangavimui atlikti.

Mokslinėje literatūroje iki šiol nebuvo analizuotas perėjimas nuo fiksuotųjų prie lanksčiųjų svorių, skaičiuojant sudėtinį rodiklį. Šio darbo naujumas informatikos mokslų srityje – ištirtas laipsniškas perėjimas ir pasiūlytas DEA metodo taikymas kartu su nauja svorių apribojimų formuluote, kurią taikant išsprendžiama DEA modelio nepakankamos skiriamosios gebos problema, užtikrinamas visų rodiklių įtraukimas į skaičiuojamą sudėtinį rodiklį ir gaunama papildoma informacija, kuri gali būti naudojama sistemos veiklai gerinti. Taip pat buvo empiriškai ištirtas vos vieną kartą švietimo duomenų analizėje taikytas hibridinio PCA-DEA modelio, kaip alternatyvos svorių apribojimams, naudojimas švietimo sistemų efektyvumui vertinti.

Praktinis šio darbo naujumas – sukurtas naujas sudėtinis rodiklis Europos šalių švietimo sistemų efektyvumui vertinti ir pasiūlyta metodika švietimo sistemų efektyvumui analizuoti. Šis darbas yra svarbus ne tik dėl švietimo tyrimų teorijos plėtojimo, bet ir dėl praktinio pritaikomumo. Per pastarąjį dešimtmetį buvo įgyvendintos naujos Europos šalių švietimo sistemų plėtros strategijos ir ES patvirtino tikslus, kurie turėtų būti pasiekti iki 2020 m. penkiose srityse, tarp kurių yra ir švietimas. Remiantis šiame darbe pasiūlyta švietimo sistemų efektyvumo vertinimo metodika, atsižvelgiant į ET2020 tikslus, galima įvertinti Europos šalių švietimo strategijų įgyvendinimo sėkmę ir nustatyti analizuojamų šalių švietimo sistemų tobulintinas sritis. Be to, pasiūlyta metodika gali būti priemonė nustatant siektinus tikslus ateityje ne tik švietimo, bet ir kitose srityse.

Ginamieji teiginiai:

1. Švietimo sistemų efektyvumo vertinimas ir sudėtinių rodiklių skaičiavimas turėtų būti atliekamas taikant DEA metodą vietoje tradicinio SAW metodo.
2. Kai rodiklių skaičius yra neproporcingas vertinamų šalių skaičiui, būtina į DEA modelį įtraukti svorių apribojimus.
3. Sudėtiniam rodikliui skaičiuoti gali būti naudojamas DEA modelis su ARI svorių apribojimais ir realizuojamas perėjimas nuo fiksuotųjų prie lanksčiųjų svorių.
4. Taikant DEA modelį su ARI tipo svorių apribojimais ir naujai pasiūlytu laisvumo parametru $\sigma \in [0;1]$ galima laipsniškai pereiti nuo fiksuotųjų prie lanksčiųjų svorių.
5. Hibridinio PCA-DEA modelio, kaip alternatyvos svorių apribojimams, skiriamoji geba yra didesnė nei DEA modelio, vertinant Europos šalių švietimo sistemų efektyvumą.

Disertacijos struktūra

Disertaciją sudaro 7 skyriai, literatūros sąrašas ir priedai. Disertacijos skyriai: *Įvadas, Sudėtinis rodiklis švietimo sistemų efektyvumui vertinti, DEA metodas efektyvumui vertinti, Svorių apribojimai DEA modelyje, Duomenų dimensijos mažinimas prieš taikant DEA modelį, Švietimo sistemų efektyvumo vertinimo metodika* ir *Pagrindiniai rezultatai ir išvados.*

Pirmajame skyriuje aprašomas tyrimo objektas, darbo aktualumas ir mokslinis naujumas, bei pristatomi darbo tikslai, uždaviniai ir ginamieji teiginiai. Antrajame skyriuje aptariama švietimo sistemų efektyvumo vertinimo svarba, pasiūloma skaičiuoti sudėtinį rodiklį Europos šalių švietimo sistemų efektyvumui vertinti ir suskaičiuojamas sudėtinis rodiklis, taikant fiksuotųjų svorių metodą. Trečiajame skyriuje pereinama prie lanksčiųjų svorių metodo taikymo indeksams skaičiuoti: pristatoma tyrimų apžvalga (DEA metodo taikymas ir lyginamoji analizė švietimo srityje) ir DEA metodas efektyvumui vertinti bei atliekama Europos šalių švietimo sistemų efektyvumo analizė, taikant DEA modelį, bei aptariama šio modelio problematika. Ketvirtajame skyriuje pristatoma pirmoji alternatyva pakankamam DEA modelio skiriamosios galios užtikrinimui – svorių apribojimų įtraukimas į DEA modelį, atliekamas fiksuotųjų svorių realizavimas DEA modelyje ir pasiūlomas naujos rūšies ribojimas, kurį taikant DEA modelyje galima palaipsniui pereiti nuo fiksuotųjų svorių prie lanksčiųjų, bei detaliai parodoma, kokia informacija gali būti gauta neefektyviai veikiančių šalių švietimo sistemų gerinimui. Penktajame skyriuje pristatoma antroji alternatyva pakankamai DEA modelio skiriamajai gebai užtikrinti – duomenų dimensijos mažinimas prieš taikant DEA modelį. Šeštajame skyriuje, remiantis atliktais tyrimais, pasiūloma švietimo sistemų vertinimo metodika. Paskutiniame skyriuje pateikiamos bendrosios darbo išvados.

Pagrindiniai rezultatai ir išvados

Nepaisant to, kad sudėtinių rodiklių skaičiavimas mokslinėje literatūroje dažnai naudojamas socialiniams ir ekonominiams reiškiniams kiekybiškai vertinti, švietimo procesams vertinti ir stebėti pasaulyje, buvo skaičiuoti vos keli sudėtiniai rodikliai, o

Lietuvos švietimo sistema šiame kontekste nebuvo vertinta nė vieną kartą. Atlikus švietimo sistemų efektyvumo vertinimo tyrimų analizę, nustatyta, kad tik nedidelė dalis tyrimų atliekama šalies lygmeniu (kai DMU yra šalis), be to, dauguma tyrimų yra orientuoti tik į pradinį ugdymą, vidurinį ugdymą ar aukštąjį mokslą, o ne į visos švietimo sistemos efektyvumo vertinimą. Šiame darbe sukurtas naujas sudėtinis rodiklis Europos šalių švietimo sistemų efektyvumui vertinti, analizuojant visą švietimo sistemą, o ne atskirus jos posistemius, ir atliekant šalies lygmens analizę.

Skaičiuojant Europos šalių švietimo sistemų sudėtinius rodiklius, formaliai yra tenkinamos mokslinėje literatūroje pateiktos rodiklių (nagrinėjamu atveju išteklių rodiklių $m = 1$ ir rezultatų rodiklių $s = 8$) ir vertinamų DMU (nagrinėjamu atveju $n = 29$) skaičiaus santykio proporcijos ($n \geq 3\dot{(}m \cdot s)$ ir $n \geq 3\dot{(}m + s)$), tačiau DEA modelio skiriamoji geba yra nepakankama sudėtiniam rodikliui skaičiuoti ir šalių švietimo sistemoms ranguoti, t. y. atlikus eksperimentinį tyrimą gauta, kad daugiau nei dviejų trečdalių šalių švietimo sistemos yra Pareto optimalios (2013 m. – 18 šalių, 2014 ir 2015 m. – 20 šalių iš 29 analizuojamų šalių) ir jų sudėtiniai rodikliai yra lygūs vienetui. Atlikus eksperimentinį tyrimą, buvo nustatyta, kad skaičiuojant Europos šalių švietimo sistemų efektyvumą, kai $n = 29$, maksimalus rezultatų rodiklių skaičius gali būti $s = 3$ arba išlaikoma rodiklių ir vertinamų šalių skaičiaus santykio proporcijos ($n \geq 6\dot{(}m + s)$ arba $n \geq 8\dot{(}m \cdot s)$), tik tada užtikrinama pakankama DEA modelio skiriamoji geba Europos šalių švietimo sistemų rangavimui atlikti.

Kadangi DEA modelio skiriamoji geba yra nepakankama, sprendžiant Europos šalių švietimo sistemų efektyvumo vertinimo uždavinį. Be to, taikant DEA modelį sudėtiniams rodikliams skaičiuoti, gauta, kad didelė dalis svorių yra lygūs nuliui, vadinasi, dalis rodiklių yra ignoruojami skaičiuojant sudėtinį rodiklį, t. y. tie rodikliai, kurių svoriai yra lygūs nuliui, nėra įtraukiami į sudėtinio rodiklio skaičiavimą. Siekiant išspręsti DEA modelio nepakankamos skiriamosios gebos problemą ir užtikrinti visų rodiklių įtraukimą į sudėtinį rodiklį, būtina į DEA modelį įtraukti svorių apribojimus.

Siekiant išskirti Pareto optimalių šalių švietimo sistemas ir Europos šalių švietimo sistemų efektyvumą vertinti pagal visų atrinktų rezultatų rodiklius ($s = 8$), į DEA modelį buvo įtraukti svorių apribojimai ir gauta, kad, SAW modeliui su vienodais absoliučiaisiais svoriais naudojant normalizuotus duomenis, rezultatas (šalių rangai) yra lygiai toks pat kaip DEA modelio su vienodais absoliučiaisiais svoriais, naudojant normalizuotus duomenis. Remiantis gautais rezultatais, galima patvirtinti galimybę naudoti DEA modelį sudėtiniam rodikliui skaičiuoti kaip alternatyvą tradiciniam SAW modeliui. Taikant DEA modelį nebūtina iš anksto žinoti rodiklių svorių (jie suskaičiuojami iš duomenų), vertinamų šalių svoriai gali skirtis (atsižvelgiama į šalies kontekstą), nėra būtina normalizuoti duomenis, be to, gaunamas ne tik šalių švietimo sistemų rangavimas, bet ir papildoma informacija, kuri gali būti naudojama vertinamų švietimo sistemų veiklai tobulinti.

Kadangi visiems rodikliams taikant vienodus absoliučiuosius svorius

DEA modelis praranda savo privalumus, buvo pereita prie ARI tipo svorių apribojimų. Atlikus empirinį tyrimą, gauta, kad taikant DEA modelį su vienodais absoliučiaisiais svoriais ir DEA modelį su ARI tipo svorių apribojimais, nepriklausančiais nuo DMU, suskaičiuoti sudėtiniai rodikliai stipriai koreliuoja (koreliacijos koeficientas lygus 0,942), vidurkiai yra panašūs (0,771 ir 0,779), o medianų skirtumas lygus nuliui (taikant Vilkoksono ženklų testą priklausomoms imtims nulinė hipotezė nėra atmetama, p reikšmė = 0,905), todėl galima pereiti prie DEA modelio su ARI svorių apribojimais ir realizuoti perėjimą nuo fiksuotųjų prie lanksčiųjų svorių (toks perėjimas mokslinėje literatūroje iki šiol nebuvo analizuotas).

Šiame darbe pasiūlyta nauja svorių apribojimų formuluotė, kurią taikant kartu su DEA modeliu yra realizuojamas laipsniškas perėjimas nuo fiksuotųjų prie lanksčiųjų svorių, išsprendžiama DEA modelio nepakankamos skiriamosios gebos problema ir užtikrinamas visų rodiklių įtraukimas į skaičiuojamą sudėtinį rodiklį. Laipsniškam perėjimui nuo fiksuotųjų prie lanksčiųjų svorių realizuoti buvo pasiūlyta nauja svorio lankstumo išraiška, kur laisvumo parametras σ kinta intervale $[0; 1]$, kai $\sigma = 1$, turime 100 proc. fiksuotuosius svorius, kai $\sigma = 0$, gauname DEA modelį be svorių apribojimų (100 proc. lankstieji svoriai), kuo σ mažesnis, tuo turime didesnį svorių apribojimų lankstumą. Darbe empiriškai pademonstruota, kad DEA modelio geba išskirti šalis didėja, didėjant laisvumo parametrui σ.

Darbe parodyta, kad taikant dalinį svorių lankstumą, gali būti pasiektas kompromisas, nustatant šalims artimiausius kaimynus ir siektinus tikslus, t. y. efektyviai veikiančios šalys (artimiausios kaimynės), kurių rodiklių reikšmės yra mažesnės nei neefektyviai veikiančios šalies, gali būti atmestos, sudarant neefektyviai veikiančios šalies siektinus tikslus. Tokiu atveju, esant daliniam svorių lankstumui, atsižvelgiama į neefektyviai veikiančios šalies kontekstą ir nustatomos būtent tai šaliai artimiausios kaimynės, o ne viena šalis visoms neefektyviai veikiančioms šalims, kaip nutinka, esant fiksuotiesiems svoriams. Be to, analizuojant švietimo sistemos rodiklius kaip šalių vidurkį, DEA modelis gali būti naudinga priemonė ES strategijai kurti ir siektiniems tikslams nustatyti, remiantis esamu šalių pasiekimų lygiu.

Ištyrus hibridinio PCA-DEA modelio (iki šiol buvo taikytas tik viename tyrime švietimo duomenims), kaip alternatyvos DEA modelio svorių apribojimams, tinkamumą Europos šalių švietimo sistemų efektyvumui vertinti, gauta, kad skiriamoji DEA modelio geba padidėjo – vietoj dviejų trečdalių efektyviai veikiančių šalių gauta, kad trečdalio šalių švietimo sistemos yra Pareto optimalios sistemos (2013 m. vietoj efektyviai veikiančių 18 šalių gauta 8 efektyviai veikiančios šalys, 2014 m. vietoj 20 šalių – 10 šalių, 2015 m. vietoj 20 šalių – 11 šalių), tačiau išliko nepakankama analizuojamų Europos šalių švietimo sistemų rangavimui atlikti. Remiantis atliktais tyrimais, buvo pasiūlyta švietimo sistemų vertinimo metodika.

Iš atliktų tyrimų galima daryti šias išvadas:

1. Skaičiuojant Europos šalių švietimo sistemų sudėtinius rodiklius, taikant DEA modelį, yra tenkinamos mokslinėje literatūroje pateiktos rodiklių ir vertinamų DMU skaičiaus santykio proporcijos, tačiau DEA modelio skiriamoji geba yra nepakankama sudėtiniam rodikliui skaičiuoti ir šalių švietimo sistemoms ranguoti. Be to, daliai rodiklių yra priskiriami nuliniai svoriai, todėl šie rodikliai nėra įtraukiami į sudėtinio rodiklio skaičiavimą. Siekiant išskirti Pareto optimalių šalių švietimo sistemas ir Europos šalių švietimo sistemų efektyvumą vertinti, atsižvelgiant į visus rodiklius, į DEA modelį turi būti įtraukti svorių apribojimai.

2. Gautas rezultatas (šalių rangai) yra identiškas, taikant SAW modelį su vienodais absoliučiaisiais svoriais ir DEA modelį su vienodais absoliučiaisiais svoriais, todėl galima patvirtinti galimybę naudoti DEA modelį sudėtiniam rodikliui skaičiuoti kaip alternatyvą tradiciniam SAW modeliui.

3. Taikant DEA modelį su vienodais absoliučiaisiais svoriais ir DEA modelį su ARI tipo svorių apribojimais, nepriklausančiais nuo DMU, suskaičiuoti sudėtiniai rodikliai yra labai panašūs, todėl galima teigti, kad sudėtiniam rodikliui skaičiuoti gali būti naudojamas DEA modelis su ARI svorių apribojimais ir realizuojamas perėjimas nuo fiksuotųjų prie lanksčiųjų svorių.

4. Taikant DEA modelį su ARI tipo svorių apribojimais ir naujai pasiūlytu laisvumo parametru $\sigma \in [0;1]$ galima laipsniškai pereiti nuo fiksuotųjų prie lanksčiųjų svorių, išs-

prendžiama DEA modelio nepakankamos skiriamosios gebos problema, užtikrinamas visų rodiklių įtraukimas į skaičiuojamą sudėtinį rodiklį ir gaunama papildoma informacija, kuri gali būti naudojama švietimo sistemai tobulinti.

5. Nepaisant to, kad hibridinio PCA-DEA modelio, kaip alternatyvos svorių apribojimams, skiriamoji geba yra didesnė nei DEA modelio, vertinant Europos šalių švietimo sistemų efektyvumą, tačiau ji išliko nepakankama analizuojamų šalių rangavimui atlikti.

Dovilė Stumbrienė

DATA ENVELOPMENT ANALYSIS IN EDUCATION SYSTEMS EFFECTIVENESS RESEARCH

Summary of a Doctoral Dissertation
Natural Sciences,
Informatics N 009

Editor Zuzana Šiušaitė

Dovilė Stumbrienė

DUOMENŲ APGAUBTIES METODAS ŠVIETIMO SISTEMŲ EFEKTYVUMO ANALIZĖJE

Daktaro disertacijos santrauka
Gamtos mokslai,
Informatika N 009

Redaktorė Lina Šiurkienė

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, 616-617 kab.,
LT-10222, Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 50 egz.