

Uncertainty and Sensitivity Analyses of PISA Efficiency: Distance Based Analysis Approach

**Marina Dobrota, Veljko Jeremić, Milica Bulajić,
Zoran Radojičić**

University of Belgrade, Faculty of Organizational Sciences,
Jove Ilića 154, 11000 Belgrade, Serbia

E-mail: dobrota.marina@fon.bg.ac.rs, jeremic.veljko@fon.bg.ac.rs,
bulajic.milica@fon.bg.ac.rs, radojicic.zoran@fon.bg.ac.rs

Abstract: The evaluation of education systems and the measurement of their quality, have become increasingly important research subjects for all stakeholders. This paper measures the efficiency of countries' educational performance and ranks countries according to the obtained results. The research is complemented by an examination of how uncertainty and sensitivity analyses affect the changes in rankings. Efficiency of countries' educational performance is presented through a single composite indicator that consists of PISA indicators and is measured using the Distance Based Analysis (DBA) methodology. Uncertainty and sensitivity of a composite indicator are results of different normalization methods. One of the contributions of this paper is the analysis of uncertainty and sensitivity impact on the efficiency measurement.

Keywords: uncertainty analysis; sensitivity analysis; composite indicators; normalization; educational performance; efficiency; PISA; Distance Based Analysis (DBA)

1 Introduction

An economic and social analysis of efficiency in education and training systems significantly contributes to the increase of education system quality. Raising the education level of the population has been recognized as a significant factor of societal economic development [1]. Although the concept of educational efficiency is proposed in several papers [2, 3] and is yet to be improved and discussed, a significant question on how stable these efficiency measurements are remains. Since it is virtually impossible to observe this issue as a one-dimensional problem, multi-criteria efficiency measurement is used and formed as a particular composite indicator. Consequently, its stability ensures an amount of safety of the observed system. Hereupon, it is essential to elaborate more on the importance of securing the safety of a complex system; the issue that has been recognized by various risk analysts in industrial and nonindustrial sectors [4, 5, 6, 7, 8, 9, 10].

The selection of an appropriate methodology is essential to any effort to capture and summarize interactions among the individual indicators included in one composite indicator or ranking system [11, 12].

According to Paruolo, Saisana and Saltelli [13], composite indicators, within social sciences, aggregate individual variables with the aim to capture relevant and possibly latent dimensions of reality. These authors claim that composite indicators have been increasingly adopted by many institutions, both for specific purposes and for providing a measurement basis for shaping broad policy debates. Composite indicators are applied and constructed everywhere [14, 15]. Saltelli et al. [16] characterize the question of composite indicators as follows: “Composite indicators tend to sit between advocacy (when they are used to draw attention to an issue) and analysis (when they are used to capture complex, multidimensional phenomena).” The topic of composite indicators has been described by the OECD [17], which defines, in six steps, why they should be created and what the goals that should thus be fulfilled are.

Many authors emphasize the need for an explicit conceptual framework for a composite indicator and the practical use of multivariate analysis prior to the aggregation of an individual indicator [12]. Authors point to the methods for assessing the robustness of the indicator using uncertainty and sensitivity analyses [12]. Sensitivity analysis is a study of how uncertainty in the models output can be apportioned to different sources of uncertainty in a models input [18]. A related practice is an ‘uncertainty analysis’ that focuses more on quantifying uncertainty in a model output. Ideally, uncertainty and sensitivity analyses should run in tandem, with an uncertainty analysis proceeding in the current practice [16]. Sensitivity analysis can serve a number of useful purposes in the economy of modeling [16]. Uncertainty quantification studies, which may include sensitivity analysis and uncertainty analysis, are essential in risk assessments. In this context, sensitivity analysis can be used to select the most relevant parameters and to reduce the number of parameters included in the risk assessment [6].

In this regard, the usual conclusion of studies, is that multi-criteria methodology definitions suffer from a ranking instability syndrome [8, 19]. Some authors offer conflict rankings as to what is “best” [20, 21]. According to Keung, Kocaguneli and Menzies [8], given different historical datasets, different sets of best ranking methods exist under various situations.

2 Instruments

In economic terms, the efficiency can be defined as the relationship between inputs and outputs while economic efficiency is increased by gain in units of output per unit of input. In relation to education, different educational outcomes can result from a variety of combinations of inputs (teachers, buildings, class size,

curriculum, etc.) [2, 61]. The problem that confronts economists and educators is how to mix the inputs in the suitable proportions to achieve the most efficient outcome [3]. Apparently, education serves many outcomes, and some of them cannot be measured by using econometric techniques of the orthodox economic theory. Historically, there have been efforts to increase educational efficiency. In order to explain the problems inherent in the measurement of educational efficiency [59], it is necessary to examine some research done on the input-output production functions. Cooze [2] for example, suggests a class size versus student achievement, which is one of the guidelines used in our research.

This paper measures the uncertainty and sensitivity of efficiency measurement of countries' educational performance. The educational performance has been described by the three main lifelong learning indicators as defined by the Program for International Student Assessment (PISA). PISA includes comparative international tests of students' achievements. It has become increasingly popular over the last ten years [22], with a major impact on the educational community, public, and national and international policy makers [23]. The relevance of knowledge and skills measured by PISA is confirmed by recent studies [23] that have been keeping track of the young people in years after they were assessed. Studies in Australia, Canada, and Denmark show a strong interconnection between a student's performance in reading and a chance to complete secondary school and carry on with post-secondary studies at the age 19 [3]. There are countries that share similarities in education, such as universal public systems and diversity in student population. Cross-cultural studies on PISA results [23, 24, 25, 26] find that some are performing at the top level while some are average performing countries [24]. According to Lounkaew [27], the differences in achievements between urban and rural students can be explained by intangible and immeasurable school characteristics. The impact of a student's family as well as school particular influences on student achievements are features that vary along the test achievement distributions [27]. However, so far PISA has been the only international assessment to incorporate the measures of multiple components in order to measure educational performance [24].

Attitudes to PISA significance differ from country to country. For example, while Germany substantially reformed its education system in a response to its average PISA results, England has made almost no change [28, 29, 30]. However, Shanghai's performance in PISA 2009 produced a global 'PISA-shock', which has repositioned this system as a significant new 'reference society' [31]. It shifted the global gaze in education from Finland to the 'East' at the beginning of the so-called 'Asian century' [31].

Measuring countries' educational efficiency is based on the set of input indicators, which envelope the appropriate financial and non-financial indicators, covering and defining the causes of a lifelong learning performance. Financial indicators are used because with a reduction of public expenditure and general rationalization policies developed countries are increasing their expenditures on

research and education [1]. At the same time, they are increasing demands for greater efficiency and effectiveness of their education system [1]. Non-financial indicators are mainly related to a class size and student-teacher ratio [2]. The input indicators used in this study are:

- Annual expenditure per student by educational institutions for all services relative to GDP per capita
- Expenditure on educational institutions as a percentage of GDP
- Ratio of students to teaching staff in the educational institution
- Starting salary/minimum training

Annual expenditure per student by educational institutions for all services relative to GDP per capita provides a measure of the cost of education. The access to education at lower levels of schooling in most OECD countries is universal. Spending per student by educational institutions in terms of GDP per capita can be interpreted as a mean of school's expenditure on the population relative to a country's ability to pay. This measure is difficult to explain on the higher levels of education, as enrolment varies from country to country. For example at the tertiary level, OECD countries can be relatively highly ranked according to this measure if a large part of their wealth is spent on education of a relatively small number of students [32, 1, 2].

Expenditure on educational institutions as a percentage of GDP is a measure of expenses on educational institutions in relation to the wealth of the country. National wealth is estimated based on GDP while the expenditure on education includes spending by governments, enterprises and individual students and their families. The share of education expenditure in GDP depends on the preferences of different public and private actors. In 2009, OECD countries spent on average 6.2% of GDP on educational institutions. In general, OECD countries spend 6.4% of GDP on educational institutions, having in mind both public and private sources of funding [32, 33].

The ratio of students to teaching staff, in educational institutions, compares the number of students to the number of teachers for a given level of education and in similar types of educational institutions. Class size and student-teacher ratios are often discussed aspects of education topics. They have a significant impact on the amount of money spent on education. Smaller departments often allow teachers to focus on the needs of a student as an individual, rather than to deal with disturbances in the department. Ratio of students to teaching staff shows how to allocate resources for education. A small student-teacher ratio is often in conflict with the possibility of higher salaries for teachers, their professional development and training, and greater investment in technology [33, 34]. It can also be in conflict with a larger number of teaching assistants whose salaries are often much lower than the salaries of qualified teachers [33, 34].

Starting salary/minimum training - teachers' salaries are the highest item of expense in school education. Together with alternative employment opportunities, salaries of teachers have a significant impact on how alluring the teaching profession is. They influence the decisions of individuals to engage in an educational profession after graduation (such as graduate career choices related to relative earnings in teaching and non-teaching jobs and their growth over time) [34]. They also impact one's decision to return and remain in the teaching profession after a career break, such as higher wages, fewer people who choose to leave the profession [34].

The countries performance (output) is made of the key competencies in reading, mathematics and science, established by PISA, adopted in 1997 by OECD countries. It represents a commitment of the governments of OECD member countries, to audit the outcomes of education systems in terms of students' achievement within a standard international framework. PISA is a collaborative effort that brings together scientific experts from the participating countries [35]. Their governments jointly steer it on the basis of shared, policy-driven interests [35]. PISA is designed to collect information through triennial assessments and presents data on domain-specific knowledge and skills in reading, mathematics and science of students, schools and countries. PISA combines the assessment of science, mathematics, and reading with information on students' home background, their approach to learning, learning environment, and their familiarity with computers. These are the background factors that student outcomes are associated with. By that, PISA provides an understanding of factors that influence the development of skills and attitudes at home and school, and examines how these factors interact and what the implications for policy development are.

The three domains assessed in PISA are:

- Reading
- Mathematics
- Science

Reading literacy presents an individual's ability to: understand, use, reflect on, and collaborate with written texts, in order to achieve goals. It allows students to develop their knowledge and potential and to participate in society [35].

Mathematical literacy is defined as an individual's ability to identify and to understand the role that mathematics has in the world, to make reasonable judgments. It allows the usage of mathematics in such a way that meets the needs of a student as a constructive, concerned, and reflective citizen [35].

Scientific literacy represents a student's scientific knowledge and its use to identify questions, acquire new knowledge, and explain scientific phenomena. It allows students to make conclusions based on scientific matters and to understand the characteristics of science as a form of human knowledge. It makes students

aware of how science and technology shape our material, intellectual, and cultural environment, and willing to engage in scientific issues [35].

The primary domains of PISA (Science, Reading and Mathematics) are a crucial part of lifelong learning. PISA indicators assess the level of performance of the adolescents. PISA provides information on teaching and learning results in schools and also demonstrates the features of the development of an educational system. The main focus of PISA is not set on the congruence of items within the national curricula of participating countries but to record the core competencies in different real-life tasks that matter in everyday situations. The concept of literacy used in PISA is functional as following: 15-year-old students should apply competencies they have learned at school in the context of authentic tasks that are part of our everyday life.

3 Methodology

Commonly, the ranking of entities is performed in a way that can vigorously affect the general evaluation issue. It can affect sports competitions, exams, medicine selection, university ranking, and many other areas [36, 37, 38, 39, 40]. For creating an objective image of ranking of the observed entities we propose I-distance method [39, 40, 41, 42].

3.1 I-distance

I-distance measures the distance in an n-dimensional space. This method has recently made a significant breakthrough in a considerable number of scientific achievements. Originally it was proposed and defined by B. Ivanovic in 1963 [37]. Affirmation of this method has been made in University ranking [39] and evaluating the socio-economic development of countries [41].

Ivanovic has initially proposed this method to rank countries according to their development level, but based on several indicators. He considered many socio-economic development indicators, but the problem was their usage in order to calculate a single indicator, which can be represented as the rank.

Let $X^T = (X_1, X_2, \dots, X_k)$ be set of variables chosen to characterize the entities.

I-distance between two entities $e_r = (x_{1r}, x_{2r}, \dots, x_{kr})$ and $e_s = (x_{1s}, x_{2s}, \dots, x_{ks})$ is defined as

$$D(r, s) = \sum_{i=1}^k \frac{|d_i(r, s)|}{\sigma_i} \prod_{j=1}^{i-1} (1 - r_{j_i, 12 \dots j-1}) \quad (1)$$

where $d_i(r, s)$ is the discriminate effect - distance between the values of a variable X_i for e_r and e_s ,

$$d_i(r, s) = x_{ir} - x_{is} \quad i \in \{1, \dots, k\} \quad (2)$$

σ_i is the standard deviation of X_i , and $r_{ji.12\dots j-1}$ is a partial coefficient of the correlation [43,44,45,46] between X_i and X_j , ($j < i$) [35, 40].

The I-distance construction is an iterative process, which can consist of several steps [47, 50]. The method is performed by calculating the mutual distances between the entities which are then compared to one another to create a rank. First step calculates the amount of discriminate effect of the first variable (the most significant variable that provides the most information on the education evaluation). Second step computes the value of the discriminate effect of the second variable, not covered by the first. This procedure is repeated for all the variables. The ranking of entities in the whole set is based on the distance from the referent entity [41, 42]. In order to perform the ranking, it is necessary to fix one entity as a referent. This is usually the entity with the minimal rate for each indicator or a fictive entity with minimal, maximal or average values for each variable. In this analysis, object with minimum values for each variable was set as a referent [47, 49].

If negative correlation coefficients and negative coefficients of partial correlations occur [48], it is more suitable to use the square I-distance, given as:

$$D^2(r, s) = \sum_{i=1}^k \frac{d_i^2(r, s)}{\sigma_i^2} \prod_{j=1}^{i-1} (1 - r_{ji.12\dots j-1}^2) \quad (3)$$

This paper also presents a Distance Based Analysis (DBA) approach as a new measure of efficiency, where the I-distance method is applied to several Input indicators ($I\text{-distance}_{input}$) and several Output indicators ($I\text{-distance}_{output}$). After I-distance values are calculated, it is necessary to normalize the values using L_∞ metrics (and other metrics). The efficiency of a country is given as follows [51, 52, 53]:

$$Ef = \frac{I\text{-distance}_{output}}{I\text{-distance}_{input}} \quad (4)$$

If the Ef value equals 1, the entity produces an equal amount of output, for given amount of input. Any country with an efficiency ratio of at least 1 is considered to be efficient [41, 48].

3.2 Sensitivity Issues

Composite indicators are filled with normative assumptions in variable selection and weighting, where ‘normative’ means ‘related to and dependent upon a system of norms and values’ [13]. Results and values of composite indicators significantly depend on a normalization method. This is why composite indicators are often the subject of controversy [54, 13].

In this paper, we propose four methods of normalization: Chebyshev distance (L_∞ norm), Manhattan distance (L_1 norm), Euclidean distance (L_2 norm), and recently proposed 1-2 distance. 1-2 distance implies the normalization of all values among values 1 and 2 thus avoiding the problems with zero values and large, poorly explainable efficiency measurement values.

Performing the sensitivity and uncertainty analyses of the PISA educational performance, in this paper we compared the results gained by these methods and pointed out the changes in rankings, in the process.

4 Sensitivity Analysis Results

In order to gain the DBA efficiency measurement values, we have calculated I-distance ranking measurement for the Input and Output variables. These are used as a basis for creating the DBA efficiency measurement of countries’ educational performance.

Input and Output values are firstly corrected with an L_∞ norm so that they could be mutually comparable, subsequently corrected with an L_1 norm, L_2 norm, and 1-2 distance. The results are given in Table 1.

When DBA method was introduced, an initially proposed normalization method was Chebyshev distance (L_∞ norm) and thus is the first to be presented in the results. Table 1 gives the values of educational performance efficiency measurement, as well as the ranking results. As can be seen from Table 1, Finland, Japan, France, Germany, Korea, Slovakia, Netherlands, New Zealand, Estonia, Australia, and Belgium are shown to be efficient in educational performance according to each normalization method. It is interesting to note that Hungary seems inefficient by Chebyshev normalization method. However, if we look at three other normalization methods, despite the fact that the values are on the verge, Hungary shows efficiency in educational performance. In addition, the Czech Republic shows even more confusing results. It is inefficient according to Chebyshev (L_∞) and Euclidean (L_2) distances, and far better positioned according to other Manhattan (L_1) normalization method. Czech Republic is even the most efficient country in educational performance according to Manhattan distance.

Table 1
Efficiency and ranking results gained by different normalization techniques

| Country | Chebyshev (L_∞) | | Manhattan (L_1) | | Euclidean (L_2) | | 1-2 | |
|----------------|--------------------------|------|---------------------|------|---------------------|------|-------|------|
| | DBA | Rank | DBA | Rank | DBA | Rank | DBA | Rank |
| Finland | 2.958 | 1 | 3.204 | 2 | 3.144 | 1 | 1.527 | 1 |
| Japan | 2.833 | 2 | 3.069 | 3 | 2.701 | 2 | 1.436 | 2 |
| France | 1.754 | 3 | 1.900 | 4 | 1.862 | 3 | 1.163 | 7 |
| Germany | 1.676 | 4 | 1.815 | 5 | 1.781 | 4 | 1.195 | 5 |
| Korea | 1.662 | 5 | 1.801 | 6 | 1.767 | 5 | 1.239 | 4 |
| Slovakia | 1.562 | 6 | 1.692 | 7 | 1.660 | 6 | 1.122 | 8 |
| Netherlands | 1.532 | 7 | 1.660 | 8 | 1.628 | 7 | 1.172 | 6 |
| New Zealand | 1.223 | 8 | 1.325 | 9 | 1.300 | 8 | 1.095 | 9 |
| Estonia | 1.219 | 9 | 1.321 | 10 | 1.296 | 9 | 1.090 | 10 |
| Australia | 1.183 | 10 | 1.282 | 11 | 1.257 | 10 | 1.079 | 11 |
| Belgium | 1.175 | 11 | 1.273 | 12 | 1.249 | 11 | 1.069 | 12 |
| Hungary | 0.945 | 12 | 1.024 | 13 | 1.004 | 12 | 1.002 | 14 |
| Switzerland | 0.910 | 13 | 0.986 | 14 | 0.968 | 13 | 0.977 | 15 |
| United Kingdom | 0.831 | 14 | 0.900 | 15 | 0.883 | 15 | 0.949 | 17 |
| Czech Republic | 0.814 | 15 | 9.357 | 1 | 0.918 | 14 | 1.360 | 3 |
| Israel | 0.808 | 16 | 0.875 | 16 | 0.859 | 16 | 1.016 | 13 |
| Ireland | 0.749 | 17 | 0.811 | 17 | 0.796 | 17 | 0.919 | 18 |
| Norway | 0.578 | 18 | 0.627 | 18 | 0.615 | 18 | 0.838 | 20 |
| United States | 0.571 | 19 | 0.619 | 19 | 0.607 | 19 | 0.833 | 21 |
| Poland | 0.526 | 20 | 0.569 | 20 | 0.559 | 20 | 0.786 | 27 |
| Spain | 0.524 | 21 | 0.568 | 21 | 0.557 | 21 | 0.849 | 19 |
| Austria | 0.513 | 22 | 0.556 | 22 | 0.546 | 22 | 0.825 | 23 |
| Sweden | 0.510 | 23 | 0.552 | 23 | 0.542 | 23 | 0.810 | 24 |
| Portugal | 0.486 | 24 | 0.526 | 24 | 0.516 | 24 | 0.806 | 25 |
| Italy | 0.450 | 25 | 0.487 | 25 | 0.478 | 25 | 0.797 | 26 |
| Chile | 0.360 | 26 | 0.390 | 26 | 0.383 | 26 | 0.956 | 16 |
| Mexico | 0.300 | 27 | 0.000 | 28 | 0.000 | 28 | 0.828 | 22 |
| Luxembourg | 0.242 | 28 | 0.262 | 27 | 0.257 | 27 | 0.621 | 28 |

Answers to this puzzle could be found in the very structure of PISA results gained by the Czech Republic. The Czech Republic scores 500 in Science, 478 in Reading, and 492 in Mathematics. When normalized, it scores 0.029 in Output according to Manhattan distance while only 0.00315 in Input. This makes Czech Republic extremely efficient according to Manhattan normalization method. Might be because L_1 metrics produces solutions that have a few large, and a lot of very insignificant residuals; the residual distribution is quite unequal (spiky).

L_2 metrics produces a few large (but less than L_1), and a lot of small, but still significant residuals. The residual distribution is more balanced than within L_1 metrics.

Table 2
Changes in Rankings

| Rankings | | | | | | | | | |
|-----------------|------------|------------|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Country | 1-3 | 4-6 | 7-9 | 10-12 | 13-15 | 16-18 | 19-21 | 22-24 | 25-28 |
| Finland | 4 | | | | | | | | |
| Japan | 4 | | | | | | | | |
| France | 1 | 2 | 1 | | | | | | |
| Germany | | 4 | | | | | | | |
| Korea | | 4 | | | | | | | |
| Slovakia | | 1 | 3 | | | | | | |
| Netherlands | | 1 | 3 | | | | | | |
| New Zealand | | | 4 | | | | | | |
| Estonia | | | 1 | 3 | | | | | |
| Australia | | | | 4 | | | | | |
| Belgium | | | | 4 | | | | | |
| Hungary | | | | 1 | 3 | | | | |
| Switzerland | | | | | 4 | | | | |
| United Kingdom | | | | | 3 | 1 | | | |
| Czech Republic | 2 | | | | 2 | | | | |
| Israel | | | | | 1 | 3 | | | |
| Ireland | | | | | | 4 | | | |
| Norway | | | | | | 3 | 1 | | |
| United States | | | | | | | 4 | | |
| Poland | | | | | | | 3 | | 1 |
| Spain | | | | | | | 4 | | |
| Austria | | | | | | | | 4 | |
| Sweden | | | | | | | | 4 | |
| Portugal | | | | | | | | 3 | 1 |
| Italy | | | | | | | | | 4 |
| Chile | | | | | | 1 | | | 3 |
| Mexico | | | | | | | | 1 | 3 |
| Luxembourg | | | | | | | | | 4 |

The uncertainty analysis, emphasizing the changes in positions and rankings, is given in Table 2 and Figure 1. Table 2 shows in detail the changes in rankings and dispersion. We can see that Finland and Japan are ranked between the 1st and the 3rd place in all four cases while France is a country with a greater dispersion of the results. Again, Germany and Korea are between the 4th and 6th place in all the

cases. It could also be noted that the dispersion is slightly increasing as we move towards lower ranking results, yet these increases are quite slight.

Figure 1 shows where the greatest differences among ranks are. If we analyze Figure 1, we can see that Manhattan and Euclidean rankings are entirely consistent with one another. As mentioned above, the largest differences are with the Czech Republic. It is ranked 15th according to Chebyshev, 14th according to Euclidean, 3rd according to 1-2, and even the 1st according to Manhattan distance normalization method. As for other countries, Chebyshev, Manhattan, and Euclidean rankings do not differ significantly from one another. On the other hand, a newly proposed 1-2 normalization method has some other differences. The most significant is Poland that is under ranked according to 1-2 norm, and Chile, which is over ranked.

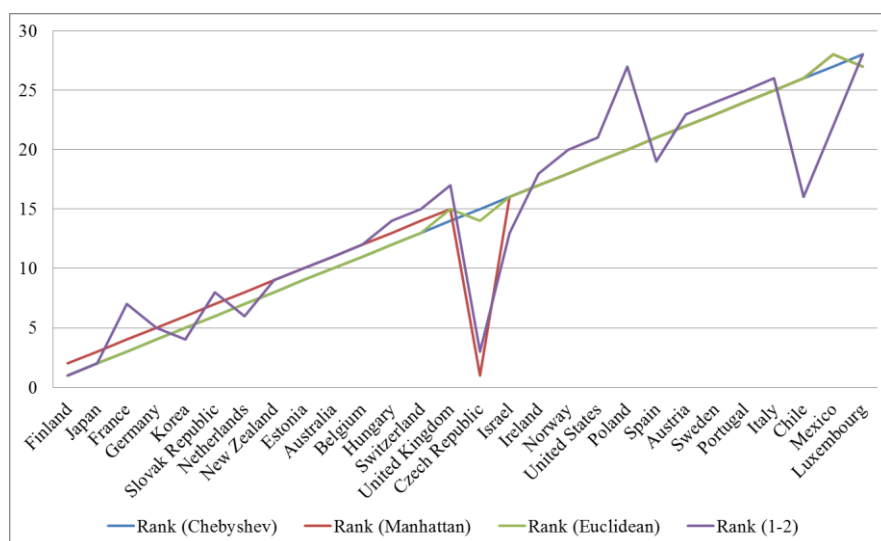


Figure 1
Ranking differences

The differences in DBA values are given in Table 1 and Figure 2.

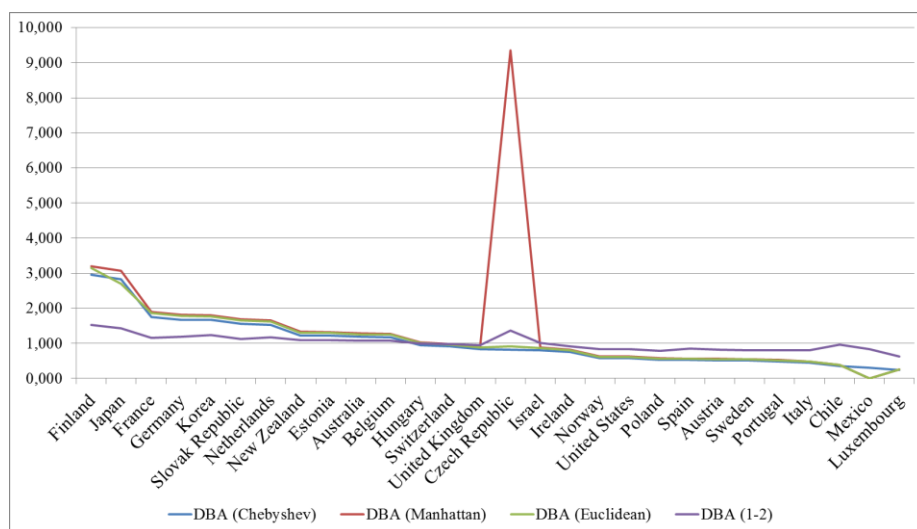


Figure 2
Changes in DBA values

Conclusions

The evaluation and measurement of certain phenomenon are not quite as simple in terms of choosing proper components to be a part of an assessment [55]. It can also be challenging, in terms of creating an appropriate methodology, which would perform the appropriate weighting or normalization [56, 57, 58]. This paper shows how uncertainty and sensitivity analyses affect the shift in rankings of countries educational performance efficiency.

There are many ways to measure educational performance and competitiveness and to evaluate countries accordingly [59, 60]. In this paper, countries educational performance is based on the PISA indicators, because it is the most exhaustive and rigorous international program to assess and explain differences in students' performance [3].

The assessments scope, nature and collective background information are decided by leading experts in participating countries and are steered jointly by governments on the basis of shared and policy-driven interests. PISA focuses on young people's ability to use their knowledge and skills to meet real-life challenges [23, 24, 26]. This orientation reflects the change in the goals and objectives of curricula themselves. The main consideration is what the students can do with what they have learned in school and not merely whether they have mastered a particular curricular content [3].

The paper has illustrated how uncertainty and sensitivity analyses, based on the changes in a normalization method, affect rankings and evaluation. Uncertainty and sensitivity analyses were applied to gain useful insight into the reliability of a

country's ranking. Changes in ranking results mainly occur due to the application of different normalization equations. While Manhattan distance measures the distance between two points in a grid based on a strictly horizontal and/or vertical path (absolute distance values), Euclidean distance uses the root of square distances (the length of the line segment connecting them). Chebyshev distance, on the other hand, is a metric defined on a vector space, where the distance between two vectors is the greatest of their differences along any coordinate dimension.

Furthermore, although the uncertainty and sensitivity analysis were performed with numerous measurement and composite indicator definitions, their impact on efficiency measurement has not yet been examined. The examination of the uncertainty influence, on efficiency, is one of the main contributions of our paper.

Acknowledgment

We are grateful for reviewers' comments and suggestions which helped us reshape and improve our paper.

References

- [1] SEEECN (2013). *Efficiency and Effectiveness of the Education System*. South East European Educational Cooperation Network, Retrieved on February 14th, 2013, from http://www.see-educoop.net/education_in/pdf/chal-educ-reform-ser-07-ser-enl-t02.pdf
- [2] Cooze, J., (1991). The Elusive Concept of Efficiency in Education. Retrieved on January 26, 2013 from <http://www.mun.ca/educ/faculty/mwatch/vol1/cooze.html>
- [3] Dobrota, M., Jeremic, V., & Martic, M. (2013). Educational Performance Efficiency of OECD Countries: PISA Lifelong Learning Indicators. In *Proceedings of the XL Symposium on Operational Research (SYM-OP-IS 2013)*, (pp. 903-908). ISBN: 978-86-7680-286-9
- [4] Arndt, S., Acion, L., Caspers, K., & Blood, P. (2013). How Reliable Are County and Regional Health Rankings? *Prevention Science*, 14(5), 497-502. doi:10.1007/s11121-012-0320-3
- [5] Monferini, A., Konstandinidou, M., Nivolianitou, Z., Weber, S., Kontogiannis, T., Kafka, P., Kay, A. M., Leva, M. C., & Demichela, M. (2013). A Compound Methodology to Assess the Impact of Human and Organizational Factors Impact on the Risk Level of Hazardous Industrial Plants. *Reliability Engineering & System Safety*, 119, 280-289. doi:10.1016/j.res.2013.04.012
- [6] Wainwright, H. M., Finsterle, S., Zhou, Q. L., & Birkholzer, J. T. (2013). Modeling the Performance of Large-Scale CO₂ Storage Systems: A Comparison of Different Sensitivity Analysis Methods. *International*

- Journal of Greenhouse Gas Control*, 17, 189-205, doi:10.1016/j.ijggc.2013.05.007
- [7] Mahsuli, M., & Haukaas, T. (2013). Sensitivity Measures for Optimal Mitigation of Risk and Reduction of Model Uncertainty. *Reliability Engineering & System Safety*, 117, 9-20. doi:10.1016/j.res.2013.03.011
- [8] Keung, J., Kocaguneli, E., & Menzies, T. (2013). Finding Conclusion Stability for Selecting the Best Effort Predictor in Software Effort Estimation, *Automated Software Engineering*, 20(4), 543-567. doi:10.1007/s10515-012-0108-5
- [9] Guttorp, P., & Kim, T. Y. (2013). Uncertainty in Ranking the Hottest Years of U.S. Surface Temperatures, *Journal of Climate*, 26(17), 6323-6328. doi:10.1175/JCLI-D-12-00760.1
- [10] Saisana, M., D'Hombres, B., & Saltelli, A. (2011). Ricketty Numbers: Volatility of University Rankings and Policy Implications. *Research policy*, 40(1), 165-177. doi:10.1016/j.respol.2010.09.003
- [11] Saisana, M., & Tarantola, S. (2002). *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*, EUR Report 20408 EN, European Commission, JRC-IPSC, Italy
- [12] Saisana, M., & D'Hombres, B. (2008). *Higher Education Rankings: Robustness Issues and Critical Assessment*. EUR23487, Joint Research Centre, Publications Office of the European Union, Italy. ISBN: 978 82 79 09704 1. doi:10.2788/92295
- [13] Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and Rankings: Voodoo or Science? *Journal of the Royal Statistical Society: Series A (Statistics in society)*, 176(3), 609-634. doi:10.1111/j.1467-985X.2012.01059.x
- [14] Brüggemann, R., & Patil, G. P. (2011). *Ranking and Prioritization for Multi-indicator Systems*, Dordrecht, Springer. ISBN:978-1-4419-8477-7
- [15] Giambona, F., & Vassallo, E. (2014). Composite Indicator of Social Inclusion for European Countries, *Social Indicators Research*, 116(1), 269-293. doi: 10.1007/s11205-013-0274-2
- [16] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Introduction to Sensitivity Analysis*. In *Global Sensitivity Analysis. The Primer*, by John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9780470725184.ch1
- [17] OECD (Nardo, M., et al.). (2008). Handbook on Constructing composite indicators –methodology and user guide (pp. 1-158). Ispra: OECD
- [18] Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd.

- [19] Jovanovic, M., Jeremic, V., Savic, G., Bulajic, M., & Martic, M. (2012). How does the Normalization of Data Affects the ARWU Ranking? *Scientometrics*, 93(2), 319-327. doi: 10.1007/s11192-012-0674-0
- [20] Shepperd, M., & Kadoda, G. (2001). Comparing Software Prediction Techniques Using Simulation. *IEEE Transactions on Software Engineering*, 27(11), 1014-1022
- [21] Myrtveit, I., Stensrud, E., & Shepperd, M. (2005). Reliability and Validity in Comparative Studies of Software Prediction Models. *IEEE Transactions on Software Engineering*, 31, 380-391
- [22] Feniger, Y., Livneh, I., & Yogev, A. (2012). Globalisation and the Politics of International Tests: the Case of Israel, *Comparative Education*, 48, 3, 323-335, doi:10.1080/03050068.2011.622539
- [23] Kankaras, M., & Moors, G. (2014). Analysis of Cross-Cultural Comparability of PISA 2009 Scores. *Journal of Cross-Cultural Psychology*, 45(3), 381-399. doi:10.1177/0022022113511297
- [24] Lin, E., & Shi, Q.M. (2014). Exploring Individual and School-Related Factors and Environmental Literacy: Comparing US and Canada Using PISA 2006. *International Journal of Science and Mathematics Education*, 12(1), 73-97. doi:10.1007/s10763-012-9396-2
- [25] Eklof, H., Pavesic, B. J., & Gronmo, L. S. (2014). A Cross-National Comparison of Reported Effort and Mathematics Performance in TIMSS Advanced. *Applied Measurement in Education*, 27(1), 31-45. doi:10.1080/08957347.2013.853070
- [26] Fernandez, A. A., & Del Valle, A. R. (2013). Educational Inequality in Costa Rica: the Gap between Students in Public and Private Schools. Analysis with the Results of the PISA International Assessment. *Cepal Review*, 111, 37-57
- [27] Lounkaew, K. (2013). Explaining Urban-Rural Differences in Educational Achievement in Thailand: Evidence from PISA Literacy Data. *Economics of Education Review*, 37, 213-225. doi:10.1016/j.econedurev.2013.09.003
- [28] Martens, K., & Niemann, D. (2013). When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US. *German Politics*, 22(3), 314-332. doi:10.1080/09644008.2013.794455
- [29] Knodel, P., Martens, K., & Niemann, D. (2013). PISA as an Ideational Roadmap for Policy Change: Exploring Germany and England in a Comparative Perspective. *Globalisation, Societies and Education*, doi: 10.1080/14767724.2012.761811, in press

- [30] Meisenberg, G., & Woodley, M. A. (2013). Are Cognitive Differences between Countries Diminishing? Evidence from TIMSS and PISA. *Intelligence*, 41(6), 808-816, doi: 10.1016/j.intell.2013.03.009
- [31] Sellar, S., & Lingard, B. (2013). Looking East: Shanghai, PISA 2009 and the Reconstitution of Reference Societies in the Global Education Policy Field. *Comparative Education*, 49(4), 464-485. doi:10.1080/03050068.2013.770943
- [32] Education at a Glance (2011). *Financial and Human Resources Invested in Education*. Retrieved on November 7th, 2013, from <http://www.oecd.org/education/skills-beyond-school/48630868.pdf>
- [33] Education at a Glance (2011). What is the Student-Teacher Ratio and How Big are Classes? Retrieved on November 7th, 2013, from <http://www.oecd.org/education/skills-beyond-school/48631144.pdf>
- [34] EAG (2012). *Education at a Glance*. Retrieved on March 15th, 2013, from <http://www.uis.unesco.org/Education/Documents/oecd-eag-2012-en.pdf>
- [35] PISA (2013). *Programme for International Student Assessment*. Retrieved on April 6th, 2013, from <http://www.oecd.org/pisa/pisaproducts/44455820.pdf>
- [36] Al-Lagilli, S., Jeremic, V., Seke, K., Jeremic, D., & Radojicic, Z. (2011). Evaluating the Health of Nations: a Libyan Perspective. *Libyan Journal of Medicine*, 6, 6021. doi:10.3402/ljm.v6i0.6021
- [37] Ivanovic, B. (1973). *A method of Establishing a List of Development Indicators*. Paris: United Nations Educational, Scientific and Cultural Organization
- [38] Ivanovic, B., & Fanchette, S. (1973). *Grouping and Ranking of 30 Countries of Sub-Saharan Africa, Two Distance-based Methods Compared*. Paris: United Nations Educational, Scientific and Cultural Organization
- [39] Jeremic, V, Bulajic, M., Martic, M., & Radojicic, Z. (2011a). A Fresh Approach to Evaluating the Academic Ranking of World Universities. *Scientometrics*, 87(3), pp. 587-596, doi:10.1007/s11192-011-0361-6
- [40] Radojicic, Z., & Jeremic, V. (2012). Quantity or Quality: What Matters More in Ranking Higher Education Institutions? *Current Science*, 103(2), 158-162
- [41] Jeremic, V., Markovic, A., & Radojicic, Z. (2011d). ICT as Crucial Component of Socio-Economic Development. *Management*, 16(60), 5-9
- [42] Jeremic, V., Bulajic, M., Martic, M., Markovic, A., Savic, G., Jeremic, D., & Radojicic, Z. (2012a). An Evaluation of European Countries Health Systems through Distance Based Analysis. *Hippokratia*, 16(2), 175-179

- [43] Fisher, R. A. (1924). The Distribution of the Partial Correlation Coefficient. *Metron*, 3(3-4), 329-332
- [44] Guilford J. P., & Fruchter B. (1973). *Fundamental Statistics in Psychology and Education*. Tokyo: McGraw-Hill Kogakusha, LTD
- [45] Rummel, R. J. (1976). *Understanding Correlation*
- [46] Baba, K.; Ritei, S., & Masaaki, S. (2004). Partial Correlation and Conditional Correlation as Measures of Conditional Independence. *Australian and New Zealand Journal of Statistics*, 46(4), 657-664. doi:10.1111/j.1467-842X.2004.00360.x
- [47] Seke, K., Petrovic, N., Jeremic, V., Vukmirovic, J., Kilibarda, B., & Martic, M. (2013). Sustainable Development and Public Health: Rating European Countries. *BMC Public Health*, 13(77)
- [48] Jeremic, V., Jovanovic-Milenkovic, M., Martic, M., & Radojicic, Z. (2013). Excellence with Leadership: the Crown Indicator of SCImago Institutions Rankings IBER Report. *El Profesional de la Informacion*, 22(5), 474-480
- [49] Dobrota, M., Jeremic, V., & Markovic, A. (2012). A New Perspective on the ICT Development Index. *Information Development*, 28(4), 271-280. doi:10.1177/0266666912446497
- [50] Ivanovic, B. (1977). *Classification Theory*. Belgrade: Institute for Industrial Economic
- [51] Coelli, T. J., Prasada Rao, D. S., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. New York: Springer
- [52] Charnes, A., Cooper W. W., & Rhodes E. L. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, 2(6), 429-444
- [53] Guan, J., Chen, K. (2012). Modeling the Relative Efficiency of National Innovation Systems. *Research Policy*, 41(1), 102-115
- [54] Saltelli, A. (2007). Composite Indicators between Analysis and Advocacy. *Social Indicators Research*, 81(1), 65-77
- [55] Kiss, G. (2012). Measuring Hungarian and Slovakian Students' IT Skills and Programming Knowledge. *Acta Polytechnica Hungarica*, 9(6), 195-210
- [56] Munda, G. (2012). Choosing Aggregation Rules for Composite Indicators. *Social Indicators Research*, 109(3), 337-354. doi:10.1007/s11205-011-9911-9
- [57] Shen, Y. J., Hermans, E., Brijs, T., & Wets, G. (2013). Data Envelopment Analysis for Composite Indicators: A Multiple Layer Model. *Social Indicators Research*, 114(2), 739-756. doi:10.1007/s11205-012-0171-0

- [58] Tate, E. (2013). Uncertainty Analysis for a Social Vulnerability Index. *Annals of the Association of American Geographers*, 103(3), 526-543. doi:10.1080/00045608.2012.700616
- [59] Kabók, J., Kis, T., Csüllög, M., & Lendák, I. (2013). Data Envelopment Analysis of Higher Education Competitiveness Indices in Europe. *Acta Polytechnica Hungarica*, 10(3), 185-201
- [60] Szoboszlai, V., Velencei, J., & Baracska, Z. (2014). Post-Experiential Education: from Knowledge to 'Knowing'. *Acta Polytechnica Hungarica*, 11(10), 235-247
- [61] Tóth, Z. E., & Jónás, T. (2014). Enhancing Student Satisfaction Based on Course Evaluations at Budapest University of Technology and Economics. *Acta Polytechnica Hungarica*, 11(6), 95-112