PIAAC ASSESSMENT: SLOVAKIA WITHIN THE OECD

Eduard Nežinský[1][2], Ivana Studená[3]

Abstract

Non-parametric DEA models are widely used in empirical investigations in educational domain. We demonstrate the merits of this approach analysing OECD PIAAC 2012 test results. Employing the basic radial CCR model, we compute a DEA-based index and expose capabilities of DEA to identify relative strong points of the subjects under evaluation. Though DEA based ranking are shown to be similar to the one derived from simple averages, the former can potentially provide insightful information from the detailed optimization results. Although in the particular dataset Japan towered over the rest of the countries and did not allow to fully demonstrate the capabilities of the approach, we find DEA a promising technique for future intertemporal analysis of PIAAC data.

Keywords

PIAAC, data envelopment analysis, DEA based index

I. Introduction

Accumulation of knowledge and skill is regarded as one of the key drivers of economic development since the birth of economics. Understanding that "the acquired and useful abilities of all the inhabitants or members of the society" present "a capital fixed and realised, as it were, in his person" can be traced as far back as Smith (1776). From the first attempt to use human capital (HK hereafter) in empirical study explaining productivity growth, a wide variety of measures have been used. Thus Mankiw et al. (1992) employ average years of schooling as a proxy for HK directly while in Hall &Jones (1999) educational level enters the labour efficiency function derived from Mincerian wage regression. Market return to education is also used in Inclusive Wealth Report which builds methodologically on Klenow &Rodriguez-Clare (2005). Barro-Lee estimates of average educational attainment used in early studies remain the most widely used HK indicator dataset. World Bank (2006, 2011) follows the residual approach where HK is included in intangible capital calculated as total wealth net of the produced and natural capital. In Jorgenson and

[1] Centre of Social and Psychological Sciences, SAS, Šancová 56, 811 05 Bratislava, Slovakia
E-mail: eduard.nezinsky@savba.sk

[2] University of Economics in Bratislava, Dolnozemská cesta 1, 85235 Bratislava, Slovakia

[3] Centre of Social and Psychological Sciences, SAS, Šancová 56, 811 05 Bratislava, Slovakia. E-mail: ivana.studena@savba.sk

Fraumeni (1992) approach, five stages of individual´s life are separated. Subsequently, HK is considered as the monetary value of the lifetime income computed.

Apart from the mere measure of educational attainment, two international OECD sponsored tests attempt to assess the *quality* of human capital. PISA tests assess student knowledge and skill whereas PIAAC zeroes in on adults. Tests results are aggregated in total score indicators, both have been massively used in empirical literature. In general, there is a trade-off – between the scope of the data needed for a measure, the sophistication of the measure, and the number of countries for which estimates currently exist.

On the eve of adults testing in Slovakia, we provide a sketch of the PIAAC – of multidimensional performance evaluation taken in applied literature. Further, we offer an empirical example to demonstrate merits of the DEA models that are often used in the domain of educational performance assessment.

We proceed by outlining PIAAC evaluation and discussing the regular way of constructing the total performance index in Section II. In Section III we introduce DEA models and DEA-based indices to derive a theoretical *best practice* frontier for PIAAC performance. Examples of the use of non-parametric approach in international comparison studies are presented. Section IV provides empirical estimation of the frontier and relative efficiency measures for selected OECD countries. Finally, Section V concludes and outlines potential methodological refinements.

II. PIAAC adult skills evaluation

Developed and organized by the Organization for Economic Cooperation and Development (OECD), Program for the International Assessment of Adult Competencies (PIAAC) is a cyclical multidimensional assessment of adults´ skills in three domains:

- Literacy (L) – defined as "*understanding, evaluating, using and engaging with written text to participate in society, to achieve one's goals and to develop one's knowledge and potential"* (OECD, 2012)
- Numeracy (N) – assesses basic math and computational skills considered fundamental for operating in everyday life: work as well as social interactions.
- Reading components (RC) concentrates on elements of reading: reading vocabulary, sentence comprehension, and basic passage comprehension. Comparability across the range

of languages in the participating countries is ensured. Designed to provide information about the literacy at the lower end of the L spectrum.

- Problem solving (PS) in technology-rich environments – defined as "*using digital technology, communication tools, and networks to acquire and evaluate information, communicate with others, and perform practical tasks.*" (OECD, 2012)

The former two competency domains (L and N) are mandatory for participating regions or countries. However, most of the countries assess PS as well. Thus, the evaluation is three-dimensional. Each subdimension (domain) result is reported either as a scale score ranging 0 – 500 or as percentages of adults reaching given performance level. PIAAC reports five proficiency levels for L and N and four levels for PS. The results undergo a thorough statistical processing including applying sample weights, trimming extreme weights, or calibration to the U.S. Census Bureau's 2010 American Community Survey population (OECD, 2013). The results reported contain additional information about the uncertainty of each statistic in the form of standard errors.

Having PIAAC results for the three domains at hand, the problem of the overall score arises. The need for a single indicator occurs in empirical cross-country comparison studies employing students (PISA) or adults (PIAAC) performance and/or as a proxy for human capital in analyses of educational systems or public spending efficiency in the regression framework. Previous works massively employed PISA results that feature similar data structure – three evaluated domains – to those of PIAAC.

Afonso & Aubyn (2006) use PISA tests results in DEA output-oriented model, the scores subsequently enter the second stage Tobit regression. Giambona et al. (2011) measure the efficiency of the European educational systems employing PISA 2006 results in DEA model. Lower and upper bounds for efficiency scores are obtained via bootstrap algorithm. Looking at the PISA scores as output variables for the 20 countries, Agasisti (2014) assesses the efficiency of public spending on education. Agasisti et al. (2019) further develop this approach combining DEA model with discrete multiple criteria evaluation. Most recently, Coco et al. (2020) investigate the effect of students´ different socioeconomic background on educational performance measured by PISA scores by means of a sophisticated slacks-based DEA model. Sometimes, results from non-parametric and semi-parametric models are compared for robustness check. Thus, Sutherland et al. (2007) collate DEA and stochastic frontier (SFA) estimates of public spending efficiency in the primary and secondary education sector. As has been demonstrated, DEA models provide a widely applied framework for educational performance evaluation. Although most of the studies utilize PISA data, structural similarity of the tests renders DEA useful in PIAAC evaluation as well. All-in assessment would contrast with some studies that only use particular PIAAC results, e.g. numeracy proficiency.

III. Construction of DEA-based index

DEA is an evaluation technique designed to assess efficiency of the economic subjects, referred to as DMUs (Decision Making Units). Each DMU is viewed as transforming a given set of inputs into outputs given the common technology. The DEA routine is aimed at determining *best practice* – 100 % efficient – subjects. The underperforming ones would be ascribed efficiency score lower than unit based on the "distance" from the efficiency frontier generated by efficient DMUs. The technology is not confined to production of goods or services, it could present any widely defined transformation process. Thus, DEA score would represent a measure of the relative efficiency.

Given the data for inputs and outputs of the set of DMUs (**X** and **Y** respectively), the basic DEA input oriented model originating from Charnes et al. (1978) takes the form of the linear program (LP).

$$\min_{\theta,\boldsymbol{\lambda}} \quad \theta \tag{1}$$

$$\text{s.t.} \quad \theta\mathbf{x_0} - \mathbf{X}\boldsymbol{\lambda} \geq \mathbf{0} \tag{2}$$

$$\mathbf{Y}\boldsymbol{\lambda} - \mathbf{y_0} \geq \mathbf{0} \tag{3}$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \tag{4}$$

where $\boldsymbol{\lambda}$ are intensity variables and subscript "0" refers to the data of the DMU under assessment. The *envelopment* program (1) – (4) operationalizes "distance" measurement (by means of the variable θ) and determining the efficiency frontier (by means of the constraints) simultaneously. The boundary is generated as linear combinations of the efficient DMUs, thus any *projection* of the inefficient DMU onto it could be interpreted as a *benchmark* and efficient subjects involved in the particular benchmark are said to be *peers* for the DMU under consideration.

The dual of the (1) – (4) is represented by optimization program

$$\max_{\mathbf{u},\mathbf{v}} \quad \mathbf{u}^{\mathrm{T}}\mathbf{y}_0 \tag{5}$$

$$\text{s.t.} \quad \mathbf{u}^{\mathrm{T}}\mathbf{Y} - \mathbf{v}^{\mathrm{T}}\mathbf{X} \leq \mathbf{1}^{\mathrm{T}} \tag{6}$$

$$\mathbf{v}^{\mathrm{T}}\mathbf{x}_0 = 1 \tag{7}$$

$$\mathbf{u}, \mathbf{v} \geq \mathbf{0}, \tag{8}$$

where **u** and **v** represent weights assigned to outputs and inputs respectively. Constraints (6) could be viewed as the transformation of the expression

$$\text{efficiency} = \frac{\textit{aggregated outputs}}{\textit{aggregated inputs}} \leq 1$$

for each DMU. Optimization (5) – (8) delivers the optimal weighting of inputs and outputs by the respective weights **v** and **u** allowing the DMU under evaluation to emphasize its strengths. To maximize its efficiency score (5), DMU puts more weight to its relatively better (lower in value) input or (higher in value) output. The constraints (6) keep the resulting score within the sensible range of 0 – 1. Alongside, weights chosen by a particular DMU are employed to evaluate the rest of the pool preventing DMUs with “poor data” be efficient. Due to duality properties, efficiency scores from LP and DP are identical in the optimum. Thus they offer two different perspectives on efficiency.

In general, DEA assesses the outputs-to-inputs ratio. However, one is often interested in the mere performance neglecting the sources or efforts put into activities. The common way of evaluating multidimensional performance is index value usually computed as a weighted sum of subindices ascribed to the constituent dimensions (taking the average is equivalent to imposing weights equal to the reciprocal of the total number). This approach assumes constant preferences across the entities under evaluation as well as constant marginal rate of substitution between the dimensions. DEA presents a more flexible approach advocated and exemplified in a number of empirical studies.

Synthetic or composite indicators have been elaborated by Cherchye (2001, 2004, 2007a, 2008) with application to macroeconomic performance, social inclusion, EU market dynamics or technology achievement and most recently van Puyenbroeck (2020). Cherchye (2007b) and Karagiannis (2016) provide a general discussion on ‘benefit-of-the-doubt’ composite indicators. Liu (2011) offered an approach without explicit inputs entering the model.

For DEA-based indices, input matrix **X** collapses to unit vector and the optimizations (1) – (4) and (5) – (8) are given output orientation. Most markedly, we normalize outputs changing the weighted sum of the latter (6) to $\mathbf{u}^{\mathbf{T}}\mathbf{y_0} = 1$.Variable and constant returns to scale render the same results. We propose that PIAAC performance of OECD countries is assessed by DEA-based composite indicator. In programs (1) – (4) and (5) – (8) performance values in three dimensions (L – literacy, N –numeracy, and P – problem solving) present three outputs of the country. All countries are endowed by the single unit input uniformly.

IV. Empirical findings

In our analysis, OECD countries act as DMUs. Raw data for the three dimensions of PIACC 2012 results are in Appendix, Table A1. Our dataset consists of 25 OECD countries leaving aside France, Spain, and Italy due to the missing data on Problem Solving.

Three main dimensions of PIAAC performance are considered output indicators. Given the fixed unit input, the DEA-based index is calculated by means of optimization (1) – (4) for each country. Thus efficiency scores computations involve solving 25 optimizations. Efficiency frontier is determined by the single DMU – Japan. The more the inefficient country falls behind Japan, the lower the resulting score from our model. These countries´ scores range from 0,857 (Chile) to 0,984 (Finland) averaging at 0,947. The radial interpretation is that on average, there is a 5,3% gap in performance with respect to Japan. Having the single efficient dominant country in this particular dataset, all the virtues of DEA cannot be demonstrated.

Table 1: Averages and DEA-based index (scores and ranks)

	avg	*ccr*	*ravg*	*rccr*
Australia	278,7	0,983	7	3
Austria	275,9	0,966	11	9
Canada	273,6	0,959	14	13
Chile	225,4	0,857	25	25
Czech Republic	277,5	0,963	8	10
Denmark	277,5	0,966	9	8
Estonia	275,7	0,948	12	16
Finland	286,9	0,984	2	2
Germany	275,0	0,963	13	10
Greece	254,3	0,874	22	23
Hungary	271,7	0,949	16	15
Ireland	266,1	0,942	19	17
Israel	259,4	0,932	21	20
Japan	292,7	1	1	1
Mexico	229,8	0,884	24	22
Netherlands	283,6	0,974	3	6
New Zealand	279,7	0,976	6	5
Norway	280,7	0,973	5	7
Poland	267,3	0,935	17	19
Korea	273,1	0,963	15	10
Slovakia	276,9	0,956	10	14
Slovenia	261,0	0,916	20	21
Sweden	282,1	0,980	4	4
Turkey	232,2	0,861	23	24
United States	266,6	0,942	18	17

Source: Own computation

In Table 1, averages across the L, N and PS PIAAC subscores are displayed (the column *avg*). Based on those values, the ranking *ravg* is generated. Likewise, *rccr* rankings are derived from CCR scores. Naturally, Japan gets top rank since it outperforms the rest of the pool in each domain. DEA-scores-based ranking is at some variance with the average-based one showing Spearman rank correlation of 0,953. Along with Estonia, there is a four places difference for Slovakia in the rankings, DEA approach being more stringent. On the other hand, Korea is favoured by DEA despite the fact that Slovakia outperforms Korea in two dimensions (L and N). This point to the problem of *weak* efficiency and *slacks* in performance which had left neglected by CCR model and should be tackled in a refined approach.

Nevertheless, some information can be extracted from the detailed CCR results. In Table 2, components of the aforementioned sum $\mathbf{u}^{\mathbf{T}}\mathbf{y}_{\mathbf{0}}=1$ are displayed. Columns L, N, and PS represent relative contributions to the total efficiency of the countries. Label VX refers to the term $\mathbf{v}^T\mathbf{x}_0$ in optimization program and is numerically identical to the reciprocal of the efficiency score.

Table 2: Contribution to efficiency (weighted data)

DMU	INPUT VX(1)	L UY(1)	N UY(2)	PS UY(3)
Australia	1,017	0	0	1
Austria	1,035	0	0	1
Canada	1,043	0	0	1
Chile	1,167	0	0	1
Czech Republic	1,039	0	0	1
Denmark	1,035	0	1	0
Estonia	1,055	0	1	0
Finland	1,016	0	1	0
Germany	1,039	0	0	1
Greece	1,144	0	1	0
Hungary	1,054	0	0	1
Ireland	1,061	0	0	1
Israel	1,073	0	0	1
Japan	1	1	1	1
Mexico	1,131	0	0	1
Netherlands	1,026	0	1	0
New Zealand	1,024	0	0	1
Norway	1,028	0	0	1
Poland	1,069	0	0	1
Korea	1,039	0	0	1
Slovakia	1,046	0	1	0
Slovenia	1,097	0	0	1
Sweden	1,021	0	0	1
Turkey	1,162	0	0	1
United States	1,061	0	0	1

Source: Own computation

A look at the PIAAC performance data reveals that apart from Japan, countries´ best performance is PS. DEA distinguishes the relative importance on the basis of the distance to the benchmark value. Thus a smaller distance to the Japan´s weakest record of 288 in N indicates more relative importance of numeracy for countries like Denmark, Estonia, Finland, Netherlands, Greece or Slovakia.

Slovakia is characterized by better than average performance in particular subdomains L, N, and PS. The average of those is clearly above the total average of PIAAC subscores. DEA model disfavoures the results that are relatively more distinct from the best country´s (Japan) subscores. Thus the CCR score based rank is down four places. Numeracy appears to be relatively the most contributive in terms of efficiency whereas in absolute terms the best performance was in problem solving domain.

V. Conclusion and further research

Benchmarking based on DEA models is considered more realistic than setting the goals at the maximum observed levels among the data. Pursuing several extremes could prove unattainable unlike linear combinations offered by DEA. The merit of this approach can fully manifest itself in the situation with more than one efficient DMUs. The particular dataset from this study only revealed a single best practice unit. Indeed, Japan outperforms the rest of countries in each subdimension. The overall dominance should be robust to any reasonable alternative method of evaluation. However, the presented approach could be helpful in some extensions and modifications. Above all, the presence of neglected slacks in performance evaluation could bias the results. The suggested solution would involve some of the slacks based assessment covering all possible sources of inefficiency. After the second wave of PIAAC assessment, intertemporal analysis employing Malmquist index of productivity can be carried out. The freshest data will help to come up with the index capturing performance change over time. As we have demonstrated, the results may differ from the ones obtained via simple averages of subscores. On top of the key DEA features, weight restrictions reflecting preferences of evaluators could be embodied in the model which would be a compromise between the score-averaging and pure free-weighting attitude. We consider DEA approach a promising technique for the result processing of the upcoming PIAAC testing in Slovakia.

Acknowledgements

The research for this paper has been supported by the national project International assessment of the key adult competencies. (PIAAC - Programme for the International Assessment of Adult Competencies) coordinated by NUCEM (Národný ústav certifikovaných meraní vzdelávania).

References

Afonso, A., & St. Aubyn, M. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modelling*, *23*(3), 476–491. https://doi.org/10.1016/J.ECONMOD.2006.02.003

Agasisti, T. (2014). The Efficiency of Public Spending on Education: an empirical comparison of EU countries. *European Journal of Education*, *49*(4), 543–557. https://doi.org/10.1111/ejed.12069

Agasisti, T., Munda, G., & Hippe, R. (2019). Measuring the efficiency of European education systems by combining Data Envelopment Analysis and Multiple-Criteria Evaluation. *Journal of Productivity Analysis*, *51*(2), 105–124. https://doi.org/10.1007/s11123-019-00549-6

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Cherchye, L. (2001). Using data envelopment analysis to assess macroeconomic policy performance. *Applied Economics*, *33*(3), 407–416.

Cherchye, L., Lovell, C. A. K., Moesen, W., & Van Puyenbroeck, T. (2007). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review*, *51*(3), 749–779.

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

Cherchye, L., Moesen, W., & Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. *JCMS: Journal of Common Market Studies*, *42*(5), 919–955.

Coco, G., Lagravinese, R., & Resce, G. (2020). *Beyond the weights: A multicriteria approach to evaluate Inequality in Education*. Dipartimento di Economia e Finanza-Università degli Studi di Bari" Aldo Moro".

Giambona, F., Vassallo, E., & Vassiliadis, E. (2011). Educational systems efficiency in European Union countries. *Studies in Educational Evaluation*, *37*(2–3), 108–122. https://doi.org/10.1016/j.stueduc.2011.05.001

Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, *114*(1), 83–116.

Jorgenson, D. W., & Fraumeni, B. M. (1992). Investment in education and US economic growth. *The Scandinavian Journal of Economics*, S51–S70.

Karagiannis, G. (2017). On aggregate composite indicators. *Journal of the Operational Research Society*, *68*(7). https://doi.org/10.1057/jors.2015.81

Klenow, P. J., & Rodriguez-Clare, A. (2005). Externalities and growth. *Handbook of Economic Growth*, *1*, 817–861.

Liu, W. B., Zhang, D. Q., Meng, W., Li, X. X., & Xu, F. (2011). A study of DEA models without explicit inputs. *Omega*, *39*(5), 472–480. https://doi.org/https://doi.org/10.1016/j.omega.2010.10.005

OECD. (2012). *Literacy, Numeracy and Problem Solving in Technology-Rich Environments:*

Framework for the OECD Survey of Adult Skills. OECD Publishing Paris.

Sutherland, D., Price, R., Joumard, I., & Nicq, C. (2007). Performance Indicators for Public Spending Efficiency in Primary and Secondary Education. OECD Economics Department Working Papers, No. 546. *OECD Publishing (NJ1)*.

Van Puyenbroeck, T., & Rogge, N. (2020). Comparing regional human development using global frontier difference indices. *Socio-Economic Planning Sciences*, *70*. https://doi.org/10.1016/j.seps.2018.10.014

World Bank. (2006). *Where is the Wealth of Nations?: Measuring Capital for the 21st Century*. World Bank.

World Bank. (2011). *The changing wealth of nations: measuring sustainable development in the new millennium*. World Bank Publications.

Appendix

Table A1: PIAAC performance of OECD countries (2012)

	(O)LIT	(O)NUM	(O)PROB	(I)INPUT
Australia	280	267	289	1
Austria	269	275	284	1
Canada	273	266	282	1
Chile	220	204	252	1
Czech Republic	274	276	283	1
Denmark	271	278	283	1
Estonia	276	273	278	1
Finland	288	284	289	1
Germany	270	272	283	1
Greece	254	252	257	1
Hungary	264	272	279	1
Ireland	267	254	277	1
Israel	255	249	274	1
Japan	296	288	294	1
Mexico	222	207	260	1
Netherlands	284	281	286	1
New Zealand	281	271	287	1
Norway	278	278	286	1
Poland	267	260	275	1
Republic of Korea	273	263	283	1
Slovak Republic	274	276	281	1
Slovenia	256	259	268	1
Sweden	279	279	288	1
Turkey	227	217	253	1
United States	270	253	277	1