NEPC

NATIONAL EDUCATION POLICY CENTER

Review of *The Efficiency Index*

Reviewed By

Clive Belfield

Queens College, CUNY

September 2014

Summary of Review

A new report that scores and ranks national education systems based on their efficiency has been receiving considerable media attention on both sides of the Atlantic. Efficiency is measured based on test scores, and resource use is analyzed in terms of teacher wages and pupil-teacher ratios. Looking across the 30 countries, the model predicts that, in order to get a 5% increase in PISA scores, teacher wages would have to go up by 14% or class sizes would have to go down by 13 students per class. But the optimal wages and class sizes for any given country may sometimes demand an increase or decrease in one or the other factor. For Switzerland, for example, the optimal teacher salary would require wages to be cut by almost half; for Indonesia, wages would have to be increased more than three-fold. For four countries, the optimal class size is estimated at fewer than two students per teacher. These extreme findings are due, in large part, to weaknesses in each of the study's three key elements: the output measure is questionable, the input measures are unclear, and the econometric method by which they are correlated does not have a straightforward economic interpretation. The report may satisfy an apparent keenness for reports that rank countries— and especially for reports that castigate low-rank countries. But it does not advance our understanding of how to make the provision of education more efficient.

Kevin Welner

Project Director

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
http://nepc.colorado.edu

Publishing Director: Alex Molnar

This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at http://greatlakescenter.org.

REVIEW OF *THE EFFICIENCY INDEX*

Clive Belfield, Queens College, CUNY

I. Introduction

The report under review here—*The Efficiency Index*—was written by Dolton, Marcenaro-Gutiérrez, and Still, under the affiliation of GEMS Education Solutions.[1] The report estimates a series of models to examine how countries' education systems differ in their technical efficiency.

The analysis in the report is based on a set of assumptions about how to measure and model efficiency. Each country's education system is measured using PISA test scores. Its efficiency is modeled using an econometric technique called *stochastic frontier analysis* (briefly described below), whereby test scores are correlated with educational inputs.

The authors score and rank 30 OECD countries by their efficiency. They attribute relative efficiency to a country's ability to manage resources across two dimensions: teacher salaries and teacher-pupil ratios. Low-ranking countries would be expected to reorganize resources along these two dimensions of teacher salaries and teacher-pupil ratios to become more efficient.

II. Findings and Conclusions of the Report

The report identifies Finland, followed by Korea, as having the most efficient education system. By contrast, "Mediterranean countries exhibit, in general, quite low efficiency" (p.16), and Brazil is the least efficient. The two leading countries "achieve good results, pay teachers reasonable wages, and have relatively high pupil/teacher ratios" (p.16). Other than these two dimensions, no other inputs or characteristics (such as private tutoring, school buildings, teacher qualifications, etc.) are considered by the report's authors to be relevant in determining efficiency and hence are not considered to be influential policy levers. Interestingly, the report takes the position that the key to efficiency is not necessarily to pay teachers more (or less) or increase (or decrease) class sizes too much, but for each country to strike a balance in resource allocation along these two dimensions, as specified by the authors' model. This balance varies from country to country. As the

authors point out, “increased spending on teachers’ salaries may not achieve the hoped for improvements in performance. . . [and] from an efficiency perspective, class sizes can be both too big and too small” (p.8).

In its conclusion, the report emphasizes that, despite their many differences, the most efficient countries are “very similar in a few key policy areas” (p.30), including teacher pay and class size. Countries can become more efficient not by spending more money on education but by allocating the money they have in a different way (and especially with respect to pay and class size). Drawing an analogy to car companies before the invention of the hybrid car, the authors then exhort education systems to shed their old practices, innovate and improve.

III. The Report’s Rationale for Its Findings and Conclusions

This report applies an econometric technique called *stochastic frontier analysis* (SFA) to estimate the efficiency of education systems. SFA generates technical efficiency scores. Systems that produce more output given their inputs have higher SFA scores. The authors then use these higher scores to rank countries in their efficiency. As part of this analysis, the report identifies two factors correlated with higher outcomes: teacher salaries and pupil-teacher ratios. The authors then simulate potential changes in resource use that would make other countries more efficient. These changes would generate technological change and modernize education systems across the world.

These arguments rely on a series of assumptions, of which three are worth highlighting.

- The primary assumption is that the outputs of a country's education system can be accurately measured by how well one age cohort of students performs on a set of standardized tests. Specifically, an education system has high output if it has a high average score on its PISA language, math and science scores administered to 15-year olds.
- A second important assumption is that SFA is the appropriate way to model efficiency and that the SFA scores represent meaningful differences in efficiency. That is, efficiency should be depicted as a frontier, with nations on the frontier performing better than nations below the frontier, who instead are wasting a lot of money on their education systems.
- A third critical assumption is that all the inputs into the education system are specified correctly. There are many inputs that help boost test scores, and all must be included in the model. That is, it must be possible to easily describe and measure the efforts that parents and families put into educating their children, as well as the contributions of teachers and other school personnel to student learning.

IV. The Report's Use of Research Literature

The report relies on existing research literature to only a very limited extent.

In three areas, the report fails to integrate its analysis with the broader literature. First, the study does not relate its findings to other studies that have analyzed the PISA data. Second, it does not compare its model specifications with other studies that have examined technical efficiency using stochastic frontier analysis.

But most importantly, the study does not incorporate any evidence that pertains to its two big findings. The literature on teacher pay, incentives, and extrinsic benefits is vast. Although it may have been difficult for these authors to summarize this literature, the report could have examined whether, e.g., research in the U.S. identifies its teachers as overpaid and research in Iceland identifies its teachers as underpaid (two apparent findings from this report). As well, the research on class size is extensive. The general conclusion from this research is that smaller class sizes boost achievement.[2] Some research does support the finding that class sizes can be too small, from an efficiency perspective, although this almost certainly depends on the grade level of the class and the baseline class size.[3] Regardless, the report does not investigate any of this evidence.

V. Review of the Report's Methods

The report's methods are questionable. Each of its three main assumptions—on outputs, inputs, and how they are modeled—is open to serious challenge.

First, the notion that the quality of an entire country's education system might be captured by PISA test scores seems implausible. There is now considerable research on the *insignificance* of achievement as a reflection of human capital: the impact of cognitive gains on economic performance is extremely modest; most interventions fail to generate long-term boosts in cognition; and educational attainment is considered as much more important for economic growth.[4] Moreover, even if achievement were important, it seems unlikely that the scores of 15-year olds would fully capture national education standards across all grade spans. The U.S., for example, allocates considerable resources to special education services to help students with disabilities; many other countries have universal preschooling, provide health screening services, or both, to make an education system more equitable rather than boost test scores. None of these features is counted in the report's output measure. Accordingly, a system efficient at producing the authors' chosen outcome may not be efficient at producing a different important outcome.

Second, the report's specification of inputs is problematic. In fact, neither of the variables that the authors emphasize is strictly an input. Teacher salaries are not inputs, they are the prices of inputs (teachers). Here, too, the analysis includes an important implicit assumption: the variable that is actually used is "teacher salaries after 15 years of experience." An alternative variable—"starting teacher wages"—is not found by the authors

to be statistically significant and so is omitted from discussion. Thus, national efficiency is determined based on the salaries of only a subset of all teachers, although the report elides this distinction in most of the text. If countries differ in the numbers of teachers with this level of experience, their efficiency levels will likely differ. Similarly, the pupil-teacher ratio is not an input; it is an indicator of how many teachers there are per student (no information is reported on what grades are included). Unfortunately, here, too, the authors use loose language: they interchange the terms class size and student-teacher ratio in their analysis. These are not the same: class size refers to the number of students in each classroom; the pupil-teacher ratio is the total number of students divided by the total number of teachers. As Adams observed:

> When a school calculates a student-teacher ratio, it is based on the total number of school instructional staff divided by the total enrollment of students. This includes all specialists, such as the librarian and the art and physical education teachers. Classes can be large, even crowded, in schools that have low student-teacher ratios overall.[5]

Also, if teachers have different teaching loads, or if class sizes vary across grade levels, the two measures will be very different from each other.[6]

The model's overall predictions are completely outside any realistic policy change.

Moreover, the report argues that other inputs do not explain efficiency. This excludes such factors as the quality of school buildings, the curriculum, and net teaching time. Importantly, it also excludes family investments in education. In countries where parents are able to provide rich educational supports, it is likely that schools will produce higher levels of achievement. Only one input that would fit in this category—private tutoring—is investigated, using correlations. Using the percentage of students in a country who receive tutoring as a variable, however, the report finds that more tutoring leads to lower scores.[7]

The third question mark relates to the model itself. One issue is model specification. The authors include more than 60 variables in their model and focus on the two that they find are statistically significant. However, many of these variables are likely to be collinear (e.g., variables measuring "feel safe" and "security"), so it is not surprising that many are not statistically significant (and that inevitably some variables will attain statistical significance). A second issue is that the model uses annual amounts: test scores in 2012 depend on resource amounts (or proxies for resource amounts) in 2012. But test scores in 2012 will surely depend on resource amounts in all the years that the student was in school. A third issue with the model, and the most important, is that of substantive significance. So, New Zealand ranks 6th in efficiency, just ahead of Slovenia. Their respective efficiency scores are 87.30% and 87.28%, a difference of 0.02%. There is no obvious way to interpret this difference or understand its economic significance. (It would not be correct, for example, to interpret it as meaning that Slovenia spends 0.02% more than New Zealand). This 0.02% gap may be economically salient. But it seems unlikely; the

ranks of 6th and 7th therefore have little meaning. More generally, it is unclear what percentage gap the authors would have us believe to be economically salient.

VI. Review of the Validity of the Findings and Conclusions

In light of the above issues, it is difficult to make a positive judgment about the validity of the report's findings and conclusions.

Overall, the idea of ranking the efficiency of a national education system based on its PISA scores is dubious. It may indeed be the case that Finland's education system is the most efficient—and that Brazil's is the least efficient—in generating high test scores with relatively limited resources. But the relevance of this for education policy in Israel or Indonesia seems to be very low. Almost certainly, policies for setting teacher salaries and determining class size are important for efficiency. But the empirical association is unlikely to be identifiable using SFA at the national level.

One way to appraise the merit of a model is in its predictions. For this SFA model, the overall predictions are completely outside any realistic policy change. For example, the model predicts that, looking across the 30 countries, in order to get a 5% increase in PISA scores, teacher wages would have to go up by 14% or class sizes would have to go down by 13 students per class. For Switzerland, the optimal teacher salary would require wages to be cut by almost half; for Indonesia, wages would have to be increased more than three-fold. For four countries, the optimal class size is estimated at less than two students per teacher. Not even the most fervent believers in small classes would contemplate that sort of change.

Of course, neither of these changes is at all feasible. Even the authors concede that they are “impractically drastic” (p.22). Even more worrisome is that the authors do not consider the efficiency implications of these predictions: is it even worth raising PISA test scores by 5%? Given the report's emphasis on efficiency, it is surprising that they do not investigate whether gains on PISA scores pass a cost-benefit test.

VII. Usefulness of the Report for Guidance of Policy and Practice

The report is founded on the suppositions that efficiency is best analyzed through international comparisons, and that this should be done using test score data. Both of these suppositions are extremely problematical and this type of analysis has contributed very little to the improvement of educational policy in the U.S. It is hard to imagine how the economic, social, and political conditions in Finland (or Korea or any other high-rank nations) necessary for educational success might be replicated in the U.S. (or in Indonesia or Brazil); and it is equally hard to imagine that U.S. citizens would welcome these potentially enormous changes simply to boost test scores at age 15. Indeed, the policy

implications that flow from the report's own analysis are unrealistic and, by the authors' own admission, "practically impossible" (p.20). The suggestion that—in order to be as efficient as Finland—some countries should have one teacher for every two students is nonsensical. But the policy implications are also paradoxical. It is extremely puzzling to understand how teacher status in the U.S.—so important to Finland's success—would be enhanced by reducing salaries by 5%, as suggested by the analysis. Similarly, it is hard to believe that more private tutoring causes test scores to go down.

Overall, the report serves to distract policymakers. The report's policy proposals derive from an abstract model that gives very little consideration to existing research (which is now extensive and of high quality), to how resources are allocated, and to how education professionals make decisions. It uses international test scores that, even if they are illustrative, cannot possibly provide a nuanced explanation of differences in human capital between the U.S. and (for example) Slovenia, if such a comparison is even worth making. And the model's predictions are far beyond what is reasonable or feasible. As such, the report may satisfy an apparent keenness for reports that rank countries—and especially for reports that castigate low-ranking countries. But it will not help advance our understanding of how to make the provision of education more efficient.

Notes and References

1 Dolton, P., Marcenaro-Gutiérrez, O., & Still, A. (2014). *The Efficiency Index*. London: GEMS Education Solutions. Retrieved September 9, 2014, from http://www.edefficiencyindex.com/book/files/assets/common/downloads/The%20Efficiency%20Index.pdf.

2 Finn, J. D., Gerber S. B. & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, *97* (3), 214–223.

3 Michel, L. & Rothstein, R. (2002). *The Class Size Debate*. Washington, DC: Economic Policy Institute.

4 Heckman, J.J. & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451-464;

Levin, H.M. (2012). More than just test scores. *Prospects*, UNESCO, *43*(2), 269-284.

5 Adams, C. (2014). Class Size Crunch: What's more vital to classroom success, a great teacher or lower class size? *Administrator Magazine*. Retrieved September 9, 2014, from http://www.scholastic.com/browse/article.jsp?id=3755248.

6 This distinction is clearly cautioned in OECD publications:

OECD (2012). *Education indicators in focus,* 2012/09 (November). Retrieved September 22, 2014, from http://www.oecd.org/edu/skills-beyond-school/EDIF%202012--N9%20FINAL.pdf.

7 Potentially, this counterintuitive result may arise because the measure of tutoring is imprecise or because families will need to tutor in countries where the public education system is inefficient.

Document Reviewed:	**The Efficiency Index**
Authors:	Peter Dolton, Olivier Marcenaro-Gutiérrez, and Adam Still
Publisher/Think Tank:	GEMS Education Solutions
Document Release Date:	September 2014
Review Date:	September 23, 2014
Reviewer:	Clive Belfield, Queens College, City University of New York
E-Mail Address:	Clive.Belfield@gmail.com
Phone Number:	917-821-9219

Suggested Citation:

Belfield, C. R. (2014). Review of "The Efficiency Index." Boulder, CO: National Education Policy Center. Retrieved [DATE] from http://nepc.colorado.edu/thinktank/review-efficiency-index.