ISSN 1995-2848
OECD Journal: Economic Studies
Volume 2009

Improving Public Spending Efficiency in Primary and Secondary Education

by
Douglas Sutherland, Robert Price and Frédéric Gonand

The authors, all members of the Economics Department at the time of writing, wish to thank Michael P. Feiner, Jørgen Elmeskov, Joaquim Oliveira Martins, Greg Wurzburg and other members of the Economics Department and participants at meetings in 2006 and 2007 of W*orking Party 1* and the *Education Policy Committee* for useful comments and discussions. The authors wish to thank Chantal Nicq for statistical assistance and Veronica Humi, Sandra Raymond and Paula Simonin for secretarial assistance. The views expressed here are those of the authors and do not necessarily represent those of the OECD.

Introduction

Influenced by the perceived link between higher levels of educational attainment and growth, the education sector has seen significant reform efforts in recent years in a number of countries. Public spending in this sector has increased on average by one-fifth in real terms over the past decade and growth in terms of spending per student has also been marked in many countries (Figure 1, upper panel); governments in the OECD area now spend on average around 3% of GDP on primary and secondary education. However, a close correspondence between the level of resources and educational outcomes is difficult to demonstrate empirically: cross-sectional evidence reveals only a weak correlation between national spending per student or teaching resources and mean pupil performance in standardised tests (Figure 1, lower panels). Extra resources devoted to education do not automatically lead to commensurate improvements in outcomes.

This paper is motivated by the need to benchmark international differences in educational performance and to assess the extent to which institutional features explain such differences. It draws together work undertaken in the Economics Department of the OECD on public spending efficiency in primary and secondary education (Gonand, 2007; Gonand *et al.*, 2007; Sutherland and Price, 2007; Sutherland, *et al.*, 2007). This work examined the extent of differences in efficiency in the primary and secondary education sector, the differences in institutional settings, and possible policy options open to government to ensure the efficient use of scarce resources. The structure of the paper is as follows: the first section describes the estimates of efficiency in the primary and secondary sector. The second section discusses and documents the policy settings in place across countries in the OECD area. This is followed by an empirical analysis of policy options that may help improve public spending efficiency in the primary and secondary education sector. A final section draws conclusions from the work.

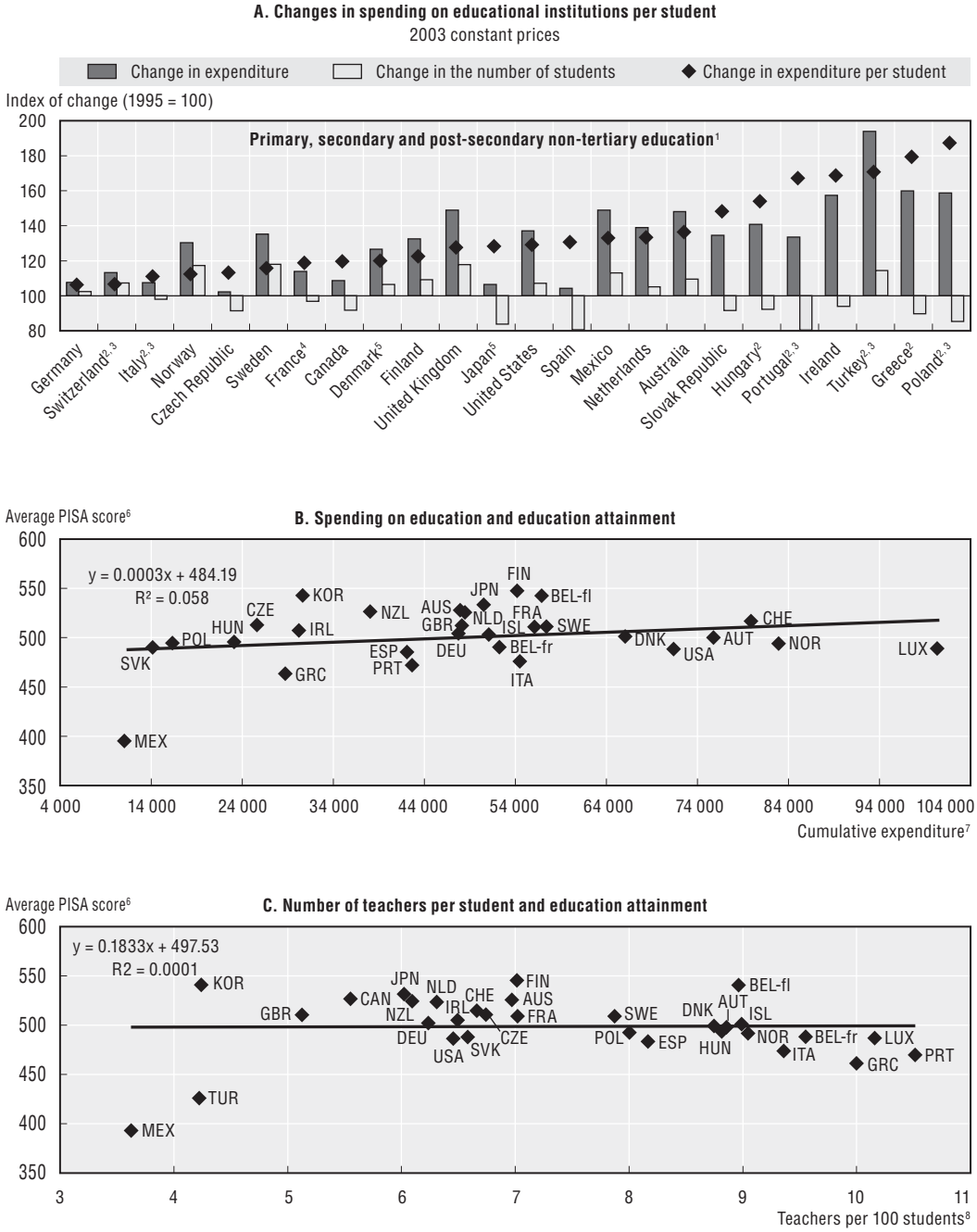
Measuring inputs and outputs in education

Measuring efficiency in the education sector is difficult due to the limitations in available data. In particular, international comparisons are hampered due to the differences in national data collection and reporting. In this light, the data used in the empirical analysis are constructed from the PISA 2003 school-level or the OECD's *Education at a Glance* national-level databases, which are both designed so that data series are comparable across countries. However, the time-series data available at the national level are relatively limited, while the richer set of school-level data are essentially limited to cross sections, which limits the analysis of efficiency to particular points in time.

Measuring inputs

Two main types of input determine educational outcomes, discretionary and non-discretionary. The first covers factors under the control of the education system. Discretionary inputs can be defined in physical inputs, such as teacher numbers, teacher-

Figure 1. **Resources and educational outcomes**



1. Countries are ranked in ascending order of change in expenditure on educational institutions per student.
2. Public expenditure only.
3. Public institutions only.
4. Year of reference 2002.
5. Post-secondary non-tertiary included in both upper secondary and tertiary education.
6. Student performance on the combined reading, scientific, mathematical and problem solving scales in 2003.
7. Cumulative actual expenditure on educational institutions per pupil between the ages of 6 and 15 converted to 2002 US dollars using PPPs for household final consumption.
8. Number of teachers per 100 students in primary and secondary education in 2003 or latest year available: 2001 for Canada, 2002 for Korea, Portugal and Switzerland.

Source: OECD (2004, 2005a, 2006); OECD calculations.

student ratios, class sizes, instruction time, teacher quality and, to a lesser extent, other resources in schools. Discretionary inputs may also be expressed in terms of spending and the national databases report spending on primary and lower secondary education. Spending differs somewhat more among OECD countries than the physical inputs, mainly due to disparities in unit labour costs. These disparities may reflect teaching quality and the availability of other potentially important resources available in schools, as well as labour-market factors unrelated to efficiency. This can make the use of spending problematic in the measurement of efficiency. The second type of non-discretionary input covers environmental inputs which are not amenable to direct control. Typically, student achievement is considered to be dependent on family and peer-group effects and also innate ability. While difficult to measure, these factors are often proxied by measures of socio-economic status, and in some cases also indicators of immigrant or language status.

For the empirical work reported in this paper, the main inputs considered are the ratio of teaching staff per 100 students, which are available from both the national and school-level databases and a measure of the socio-economic background of the student. The larger *PISA* dataset at the school level allows a wider set of inputs to be considered in the estimates, which include a measure of computer availability as a proxy for capital inputs and an index of whether the language spoken at home is a national language to assess potential barriers facing children of immigrants and minority groups in education.[1]

Measuring outputs

Educational output, at its most basic level, can be measured by "quantity" indicators such as course enrolment and completion rates, study duration, the level of education reached, or even equated with the quantity of inputs. However, the narrow nature of academic exams has been criticised in that it may not link directly to welfare objectives or growth. An approach which takes the quality of teaching (and learning) into account would focus more on outcomes, such as literacy rates at particular ages, learning achievement, or longer-term earnings and occupation would give a better understanding of educational contributions to human capital.

For the empirical work reported in this paper, the main output is a synthetic indicator of the average country and school-level PISA scores derived from principal component analysis.[2] The use of PISA results for 15 year old students gives a measure of the cumulative output of primary and lower-secondary education goes some way to treating educational outcomes as outputs. The PISA questions are designed to evaluate real-life aptitudes and not just academic attainment. Assessment at the age of 15 is designed to capture student abilities at the end of their compulsory schooling, thus avoiding the difficulties of comparing students across countries when participation is voluntary.

The use of PISA results as a measure of educational output does have some drawbacks, however. National authorities may be targeting other outputs that are more difficult to take into account in the analysis. In particular, non-cognitive skills can be difficult to measure and particularly so in an international context. As such, differences in efficiency based on a narrower set of outputs may be biased to the extent that the unmeasured outputs differ across countries and are uncorrelated with other outputs. Secondly, in a small number of OECD countries enrolment rates are relatively low at the age the PISA tests are administered (under 60% in Mexico and Turkey compared with an OECD average of 95%). Cross-country comparisons could also be distorted by high truancy rates in some countries. In these cases, PISA mean scores may tend to overestimate the average level of

human capital for the overall cohort. Related to these concerns is the fact that educational outcomes could also be judged by how many of the young continue their education to the end of the compulsory stage, which has obvious implications for human capital accumulation.

Measuring efficiency

The fact that outputs in the public sector are amorphous and intangible in many respects makes it difficult to define a supply function in the conventional sense, while the fact that public sector organisations produce goods that are free at the point of use means that the prices of outputs are not determined by market forces. As economic efficiency cannot be directly measured, a technique is needed to proxy an efficiency frontier which would allow relatively accurate benchmarking. This section looks at two main approaches for achieving this: the first is a non-parametric technique and the second stochastic frontier analysis.

The measurement of efficiency in this study relies principally on Data Envelopment Analysis, which is a non-parametric approach to measuring performance relative to a measure of best practice or inferred best practice (Box 1). In essence, this technique assesses performance relative to estimates of the additional output or the resource savings that are possible if the country or school increased efficiency to the level of best practice. Inputs are described both in physical terms, the relationship between input volumes (for example teachers) and outputs being a measure of technical efficiency, and in terms of the amount spent per pupil, which determines cost efficiency. A considerable attraction of non-parametric approaches to estimating efficiency is the relative simplicity of the procedure. For example, non-parametric approaches do not require assumptions about the *specific* functional form of the educational production function, although they do assume that the production function is common to all units. In addition, the framework can easily accommodate multiple outputs, which is an attractive feature for public services where equity objectives may be important.

The approach also has a number of drawbacks that are often aggravated in small samples or when the number of inputs and outputs rises. First, the technique is sensitive to measurement error and statistical noise. For example, an observation that erroneously determines a segment of the efficiency frontier will in turn affect the measures of inefficiency of all the schools or countries lying within that segment. Second, small samples – by frequently excluding best practice – can lead to a bias of under-estimating inefficiency. Third, omitted or irrelevant inputs will tend to over (or under) estimate the "true" degree of inefficiency. In order to counter these problems, the estimates correct for possible small-sample bias and generate confidence intervals surrounding the point estimates.

In order to test the robustness of the approach, estimates of efficiency were also made using the parametric technique stochastic frontier analysis (SFA). By making assumptions about the distribution of inefficiency it is possible to decompose the regression error term into statistical noise and a measure of inefficiency (Box 2). Consequently, SFA is less sensitive to the influence of measurement error. However, this feature comes at the cost of having to specify the functional form and making assumptions about the distribution of inefficiency.

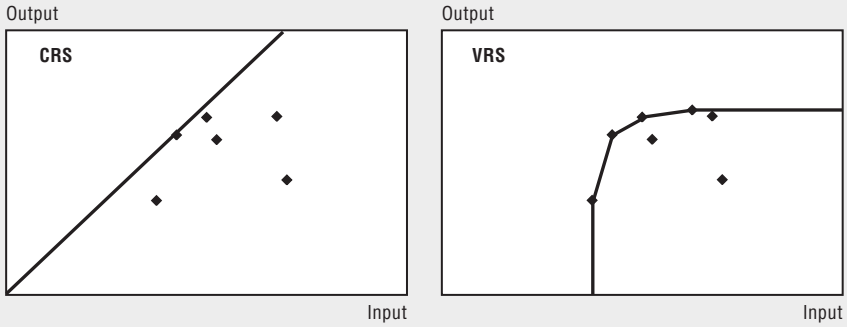
Box 1. **Data Envelopment Analysis**

Data Envelopment Analysis (DEA) constructs an efficiency frontier based on the input and output data from all the countries/schools of a sample. In essence, the frontier is constructed from the schools or countries that ***envelop*** the remaining observations and thus provides a benchmark by which the others can be judged. By assumption, the frontier determines best practice, and potential efficiency gains for specific countries or schools are measured by their position relative to the frontier or the envelope. In the "one input-one output" case shown below, a measure of the efficiency shortfall in terms of unachieved output is given by the ratio of a school's output to the output on the frontier for the same level of inputs (*i.e.* the point on the frontier vertically above the school/country observation). Conversely, the ratio of inputs on the frontier to the school's inputs at the same output (measured horizontally) is a measure of inefficiency in terms of potentially excess inputs. In the case of multiple inputs or outputs, the measures of efficiency are determined in a similar fashion by holding the relative proportions of either inputs or outputs constant in measuring the distance to the frontier.

The shape of the DEA efficiency frontier depends on the assumptions about returns to scale:

- *Constant returns to scale (CRS)*. This assumption describes the efficiency frontier as a ray from the origin through the observation(s) with the highest output/input ratio (Box figure, left panel).
- *Variable returns to scale (VRS)*. This approach identifies the schools or countries that define the frontier by starting from the observations of units that use the least of each input and ending with the observations producing the highest amount of each output (right panel).
- *Non-increasing returns to scale (NIRS)*. This assumption combines the constant returns to scale assumption between the origin and the observation with the highest output/input ratio, and variable returns to scale thereafter.

DEA efficiency frontiers



As noted in the main text, the non-parametric approach is not without drawbacks. The impact of measurement error, statistical noise and outliers that define the efficiency frontier will influence the estimates of efficiency for other observations within the frontier. Even though exploratory data analysis should help eliminate observations that are obvious outliers, the results are still sensitive to measurement error and statistical noise. Second, in a finite sample, particularly when the number of inputs and outputs rises, the efficiency estimate will be biased towards smaller estimates of inefficiency. Bootstrapping can help address these problems by producing confidence intervals around the point estimates as well as making a correction for the small sample bias. The bootstrap is based on the assumption

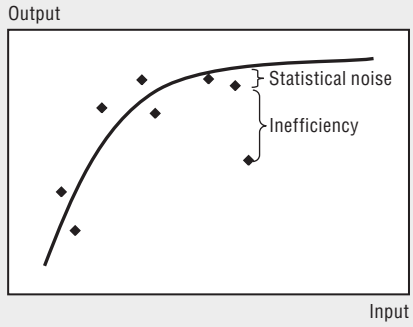
Box 1. **Data Envelopment Analysis** *(cont.)*

that the observed data are representative of the population. Taking repeated samples that are the same size as observed data will mimic sampling from the total population. Roughly speaking, repeating this sampling generates different efficiency estimates and when the number of (re)-samples is large the standard errors of these estimates can be used to derive the confidence intervals. In addition, the bootstrapping procedure allows the small-sample bias to be estimated and a correction to be made to the efficiency estimate (Simar and Wilson, 2007).

Box 2. **Stochastic frontier analysis**

Stochastic frontier analysis is similar to standard regression techniques but differs by exploiting the one-sided nature of inefficiency to decompose the error term into a standard error term and an asymmetric component that measures inefficiency as shown below. The principal advantage of this approach to measuring efficiency is that it addresses statistical noise explicitly. Furthermore, and in contrast to non-parametric approaches, standard statistical tests can be used to assess variables. The cost of specifying a specific functional form can be mitigated somewhat by the choice of a translog production function. This is a very flexible specification, also referred to as a second order approximation of an unknown functional form. However, estimating this production function in practice is often complicated by multicollinearity (the variables are too highly correlated with one another to allow their individual effects to be estimated with precision). In addition, the specification may not make economic sense for all observations and thus need to be checked. For example, the estimated coefficients may imply for a particular observation that increasing inputs would lead to fewer outputs.

Stochastic frontier analysis



School-level results

As a first step, stochastic frontier analysis was conducted using the data and specification described above. A translog production function was estimated as this production function allows greater flexibility, including non-constant returns to scale and varying input elasticities across schools.[3] The basic results are presented in Table 1, which presents the results for two specifications with different assumptions about the distribution of inefficiency (half-normal and exponential). The results suggest a significant role is played by student socio-economic background in determining performance in the PISA tests. Student performance also increases with teacher numbers, though with

Table 1. **Stochastic frontier specification**

	Half normal distribution		Exponential distribution	
	Coefficient	z	Coefficient	z
Constant	5.064	205.54	5.041	209.97
Teachers student ratio	0.044	2.91	0.041	2.77
Computer availability	–0.008	–1.26	–0.007	–1.14
Socio-economic background	0.500	74.05	0.498	75.78
Language background	0.109	15.88	0.108	15.95
Teachers student ratio * teachers student ratio	–0.055	–6.93	–0.051	–6.50
Computer availability * computer availability	–0.012	–6.22	–0.011	–6.10
Teachers student ratio * computer availability	0.021	6.18	0.019	5.90
Lambda	1.958		0.990	

Note: The table presents the estimated coefficients and associated 2-scores for trans-log production functions, assuming different inefficiency distributions. A positive coefficient signifies a positive relationship with average performance in PISA tests.

diminishing returns. Capital inputs – as proxied by computer availability – make a positive contribution to student performance only in conjunction with teaching inputs. The efficiency estimates for the median school in each country and a measure of the variation of efficiency across schools within countries are presented in Table 2.

For the second step, the same inputs and output were used in Data Envelopment Analysis. These estimates are highly correlated with the SFA estimates of efficiency. The main results of the school-level DEA estimates of technical efficiency are:

- There are significant potential gains from eliminating inefficiency. The scope to reduce inputs while holding outputs constant is around one-third for the median school (Figure 2). Potential gains from maximising outputs from the current level of inputs are proportionally slightly smaller. The average PISA score of students in the median school is around one-fifth below the level suggested possible by the efficiency frontier.[4]
- There are significant differences across countries (Figure 2), appearing to be most substantial for Norway, Iceland, Greece, Luxembourg, Hungary and the United States. The potential to eliminate inefficiency in the median schools in Korea and Japan appears comparatively modest.
- There are significant differences in the estimates of technical inefficiency ***within*** some countries, though not generally. Figure 3 computes the reduction in staffing levels and increases in PISA scores if all schools were to attain at least the degree of technical efficiency of the school at the 95th percentile. The potential gains are relatively modest at around 10%. However, the potential for resource savings compared with national good practice is greater in Belgium, Greece and Italy, while the gains in PISA scores are potentially substantial in the Czech Republic and Turkey.
- The estimates of technical inefficiency are generally much larger with the constant returns to scale assumption than the other assumptions, which is consistent with most schools being clustered along the portion of the efficiency frontier characterised by diminishing returns to scale. In the majority of cases, eliminating inefficiency by achieving the maximum possible outputs from available inputs would locate schools on this portion of the efficiency frontier. However, scaling back inputs to eliminate technical inefficiency would leave a significant portion of schools operating below their most efficient scale in Turkey, Greece, and Hungary.

Table 2. **Stochastic frontier estimates of technical efficiency at the school-level**
Preferred specification[1]

	Exponential		Half-normal	
	Median	Range 90%	Median	Range 90%
Australia	0.96	0.05	0.94	0.08
Austria	0.95	0.15	0.93	0.17
Belgium	0.97	0.11	0.95	0.14
Belgium Flemish c.	0.97	0.13	0.95	0.15
Belgium French c.	0.97	0.11	0.95	0.13
Canada	0.95	0.11	0.92	0.13
Czech Republic	0.96	0.09	0.93	0.12
Denmark	0.95	0.08	0.92	0.10
Finland	0.97	0.04	0.95	0.06
Germany	0.96	0.11	0.93	0.13
Greece	0.93	0.23	0.90	0.22
Hungary	0.95	0.21	0.92	0.21
Iceland	0.93	0.12	0.89	0.13
Ireland	0.96	0.05	0.94	0.07
Italy	0.96	0.17	0.94	0.18
Japan	0.97	0.09	0.96	0.13
Korea	0.97	0.05	0.96	0.09
Luxembourg	0.95	0.07	0.91	0.08
Netherlands	0.96	0.08	0.94	0.11
New Zealand	0.96	0.07	0.94	0.10
Norway	0.91	0.13	0.87	0.13
Poland	0.96	0.05	0.94	0.08
Portugal	0.97	0.09	0.95	0.12
Slovak Republic	0.95	0.10	0.93	0.12
Spain	0.94	0.11	0.94	0.10
Sweden	0.95	0.06	0.92	0.09
Switzerland	0.96	0.09	0.94	0.12
Turkey	0.96	0.15	0.93	0.17
United Kingdom	0.96	0.07	0.93	0.10
United States	0.93	0.18	0.89	0.17
Standard deviation	0.01	0.05	0.02	0.04
Median	0.96	0.10	0.94	0.12
Average	0.95	0.11	0.93	0.12

Note: The table gives the estimate of the level of efficiency for the median school and the range that encompasses 90% of schools in a particular country.

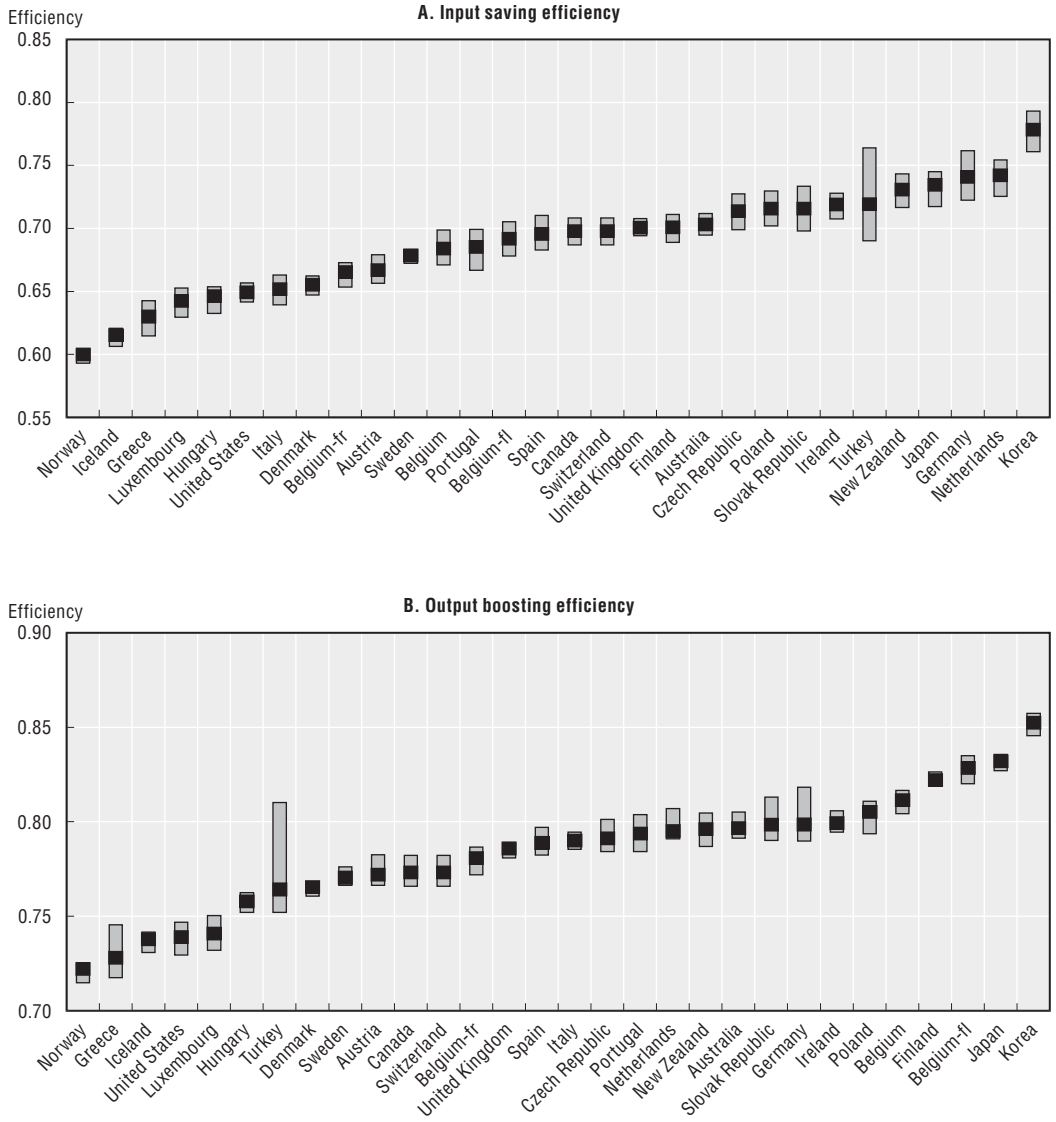
1. PISA score as output and teachers student ratio, computer availability, socio-economic and language backgrounds as inputs.

Country-level results

Estimates at the country level allow additional outputs to be considered as well as considering estimates of cost efficiency in addition to the technical efficiency.[5] However, the small-sample bias is likely to be more pronounced, particularly when there are few countries with similar input and output mixes, making estimates of the portion of the efficiency frontier that is relevant for these observations less certain. In this light, the baseline specification for the national sample restricted the inputs considered to the ratio of students to teachers and the measure of the socio-economic background of the student. The main results of the estimates of technical efficiency at the national level are:

- The potential for efficiency gains is around 19% on average and largest in Iceland, Norway and the United States (Figure 4, Panel A).[6] The potential gains to be made by

Figure 2. **Estimates of technical efficiency for the median school**

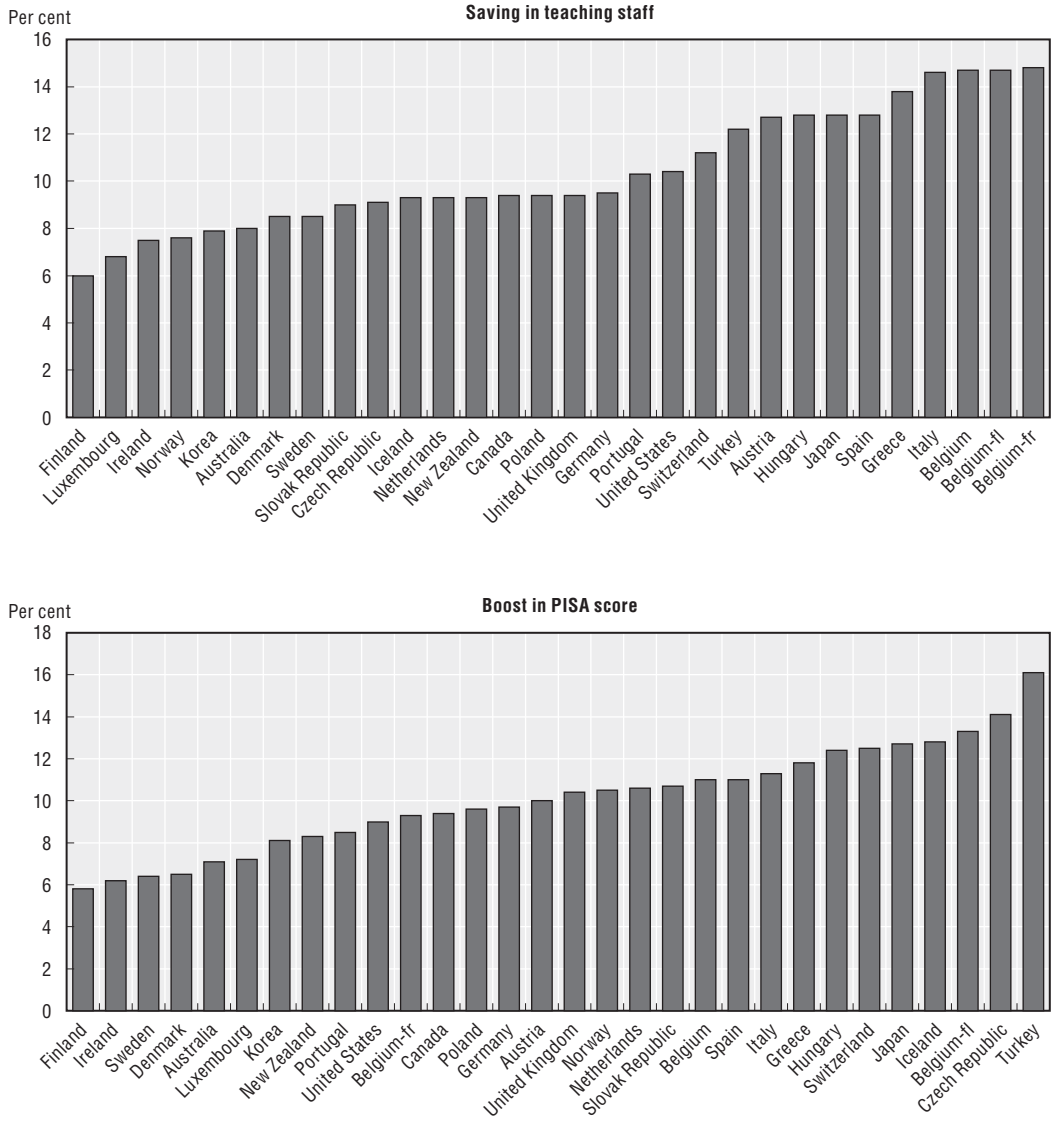


Note: DEA performed with 4 inputs (teaching and computing resources, socio-economic status of students and language background) and one output (average PISA score). The figure shows the bias-corrected estimate of efficiency and 95% confidence intervals.

eliminating inefficiency imply substantial possible improvements in performance or, alternatively, important resource savings. For example, the resource gains which would accrue if teacher-student ratios were reduced while holding outputs constant would largely be in the 0.2-0.4% of GDP range, but in some cases, notably Norway, savings could be substantially greater The scope for improving average PISA scores while holding inputs constant is much smaller: on average about 5% relative to the efficiency frontier, rising to around 10% for the least efficient (Figure 4, panels B), though this is in part driven by the assumption about the shape of the frontier.

- At the national level technical efficiency measures are broadly similar for the baseline case and the wider specification when considering how much inputs could be cut while holding outputs constant. There are some differences in ranking when the equity

Figure 3. **Raising efficiency towards national best practice**

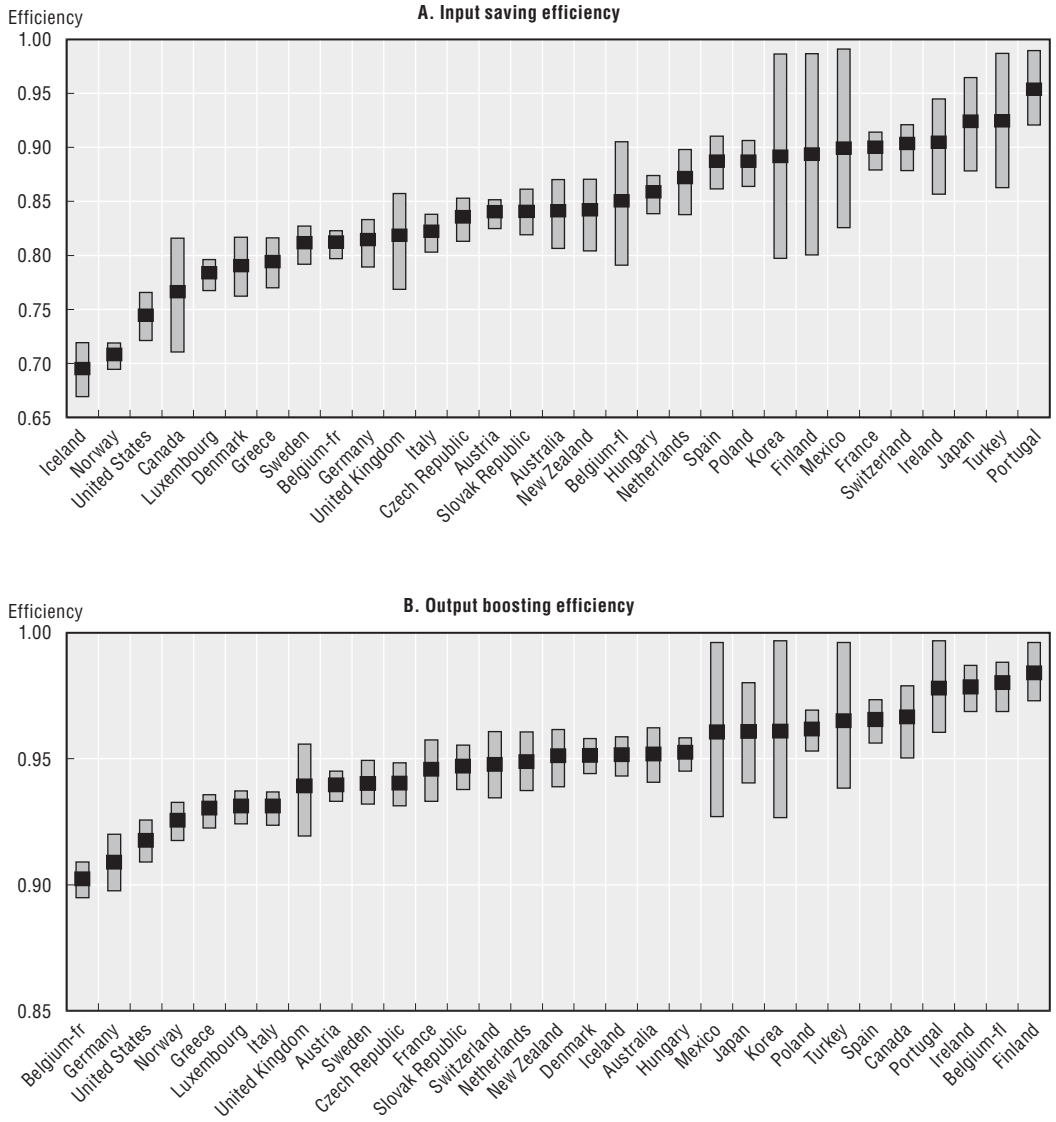


objective is also considered, though broadly speaking, countries that are inefficient when considering one output also tend to be relatively inefficient when considering two outputs. In general, countries which are the least efficient in terms of input efficiency also tend to be rather inefficient in terms of output efficiency.

- Some countries, particularly Ireland, are relatively more efficient at ensuring that educational attainment is evenly spread rather than in securing high levels of human capital alone. On the other hand, other countries, such as Germany, appear relatively less efficient when the equity objective is included as an output in the specification.[7] Taking enrolment rates into account as a third output to assess broader equity concerns generally has a relatively modest effect on the estimates of efficiency.

The most notable difference between these results and the school-level results is that estimates of inefficiency are generally much smaller at the country level, a result which is the natural outcome of the DEA method given the greater variation in the school data. The

Figure 4. **Estimates of technical efficiency at the national level**



Note: DEA performed with 2 inputs (teachers per 100 students and pupils' socio-economic background) and 2 outputs (average PISA scores and equity objective). The figure shows the bias-corrected estimate of efficiency and 95% confidence intervals.

best schools will be more efficient than the average, which will generate larger estimates of inefficiency, whereas, at the national level, aggregation masks the within-country variation. The uncertainty surrounding the efficiency estimates is often considerable and much larger than that surrounding school-level results, particularly for countries that are close to or define the efficiency frontier. However, the relative rankings are broadly consistent with the national results.

In a small number of countries, notably Mexico, Turkey and, to a lesser extent, Portugal, estimated efficiency is relatively high – partly as a result of few other countries with similar input and output mixes – but the average performance in the PISA tests is comparatively poor. A comparison of country- and school-level estimates of technical efficiency makes clear that the high efficiency scores from the country-level analysis are an

artefact of the small sample size. Portugal and Turkey appear to be among the most efficient when comparing national averages but at the school level the efficiency of the median school is substantially worse relative to other countries. The comparison of schools can identify different schools both within Portugal and Turkey and across the rest of the sample that have similar input mixes that are also more efficient, whereas within the country sample there are few other countries with similar input mixes. This suggests that caution is required in assessing the relative efficiency using aggregate country-level data for those countries close to or defining the frontier.

Estimates of cost-efficiency at the national level use almost the same baseline and wider specification as those of technical efficiency at the national level; though the teacher-pupil variable is replaced with cumulative spending. These estimates are also much more uncertain as a result of the difficulties in estimating actual spending on students and the choice of converting estimated spending into a common currency (Box 3). Bearing in mind the caveats, the main results from the estimates of cost efficiency are:

- Overall, the potential gains from eliminating cost inefficiencies are similar to those from eliminating technical inefficiency: inputs could be cut by one-sixth on average, with potential cuts of over one-quarter in Iceland, Luxembourg, Norway and the United States. Output improvements from the same level of inputs reach a maximum of approximately 15% (Figure 5).
- For some economies (notably the Czech Republic, Poland, the Slovak Republic and, to a lesser extent, Hungary), potential efficiency gains are noticeably lower than they are in terms of technical efficiency. This may be a reflection of teaching compensation (or quality) relative to other countries.
- As in the case of technical efficiency, some countries experience a relative deterioration in their performance when the outputs include the achievement of homogeneity in pupil test results. However, in general, the worst performers with respect to average PISA scores also generally emerge less favourably *vis-à-vis* the homogeneity of results.
- As with the estimates of technical efficiency at the national level, comparing cost efficiency estimates with PISA scores reveals only a weak tendency for better PISA performance to be related to better cost performance. This is not uniformly the case, however, and in some cases, notably Mexico and to a lesser extent Portugal, efficiency appears to be relatively high but PISA scores low by international comparison.

Institutional factors influencing efficiency

The data limitations inhibiting the measurement of policy on a consistent basis across countries are arguably more severe than those affecting the measurement of efficiency. Comparable cross-country data exist for only a limited number of measures of national policy that can be construed as affecting efficiency. To expand the information set an *ad hoc* questionnaire was used to collect data at the national level and the responses were used to construct composite indicators of national policy. This section describes these indicators.

Work undertaken by Gonand *et al.* (2007) established a framework for constructing composite indicators that assess the extent to which institutional and policy settings are well-designed, insofar as they display the characteristics which would be expected, *a priori*, to enhance public sector efficiency.[8] This work identified policies with efficiency-influencing characteristics in three broad policy groupings – resource allocation, budget management and market framework (described below). The three broad policy groupings

Box 3. **Comparing spending across countries**

Assessing cost efficiency is in practice more complicated than the analysis of technical efficiency. Both in constructing the relevant indicator of spending and comparing spending data across countries, particular care is needed in measuring inputs:

- Cost efficiency depends on cumulative spending on education per student between the ages of 6 and 15, this being calculated from total spending per full time equivalent student between 1993 and 2002, taking into account the progress of the student through pre-primary, primary and lower secondary education, depending on the education system of the country under consideration.
- The so-called "Balassa-Samuelson effect" suggests that the relative price of non-tradable goods (such as many services) tends to rise with per capita income. Failure to take account of this phenomenon will thus exaggerate the differences in spending between richer and poorer countries. Here, the estimates of cumulative education spending per student have been converted into constant prices using the private consumption deflator and then into constant 2002 dollars using the PPP for private consumption, which may not eliminate this possible bias.

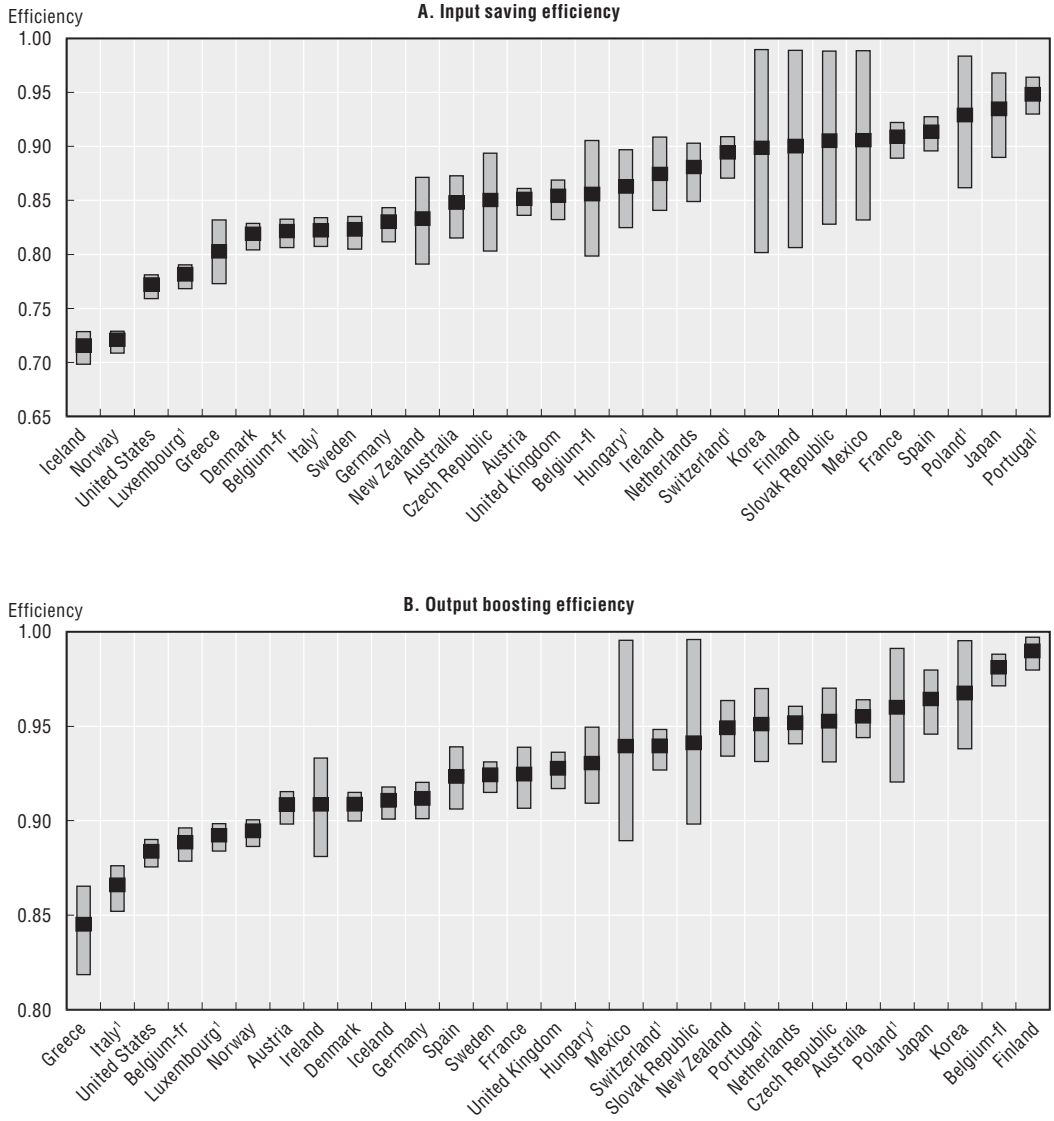
In order to check the robustness of the purchasing power parity (PPP) used in the analysis, additional estimates of cost efficiency were made (using spending on pupils in primary and secondary schools in a single year) converted into constant dollar prices using PPPs for the service sector and total output. The comparison of cost efficiency across countries is complicated by differences in relative prices in richer and poorer countries. In this regard, the differences between high and lower income countries may be exaggerated when evaluated at market exchange rates or even when using PPPs that include prices for tradable goods.

The translation of spending on education for each country into a common currency could use several different PPPs.

- The estimates of efficiency can be somewhat sensitive to the choice of deflator for cross-country comparison for some counties. For example, estimates of efficiency for the Czech Republic swing markedly when different PPPs are used. Nevertheless, the estimates of efficiency are relatively stable in most countries.
- PPPs that are education specific or relate to general government consumption (*i.e.* mainly education and health care) would be preferable. However, for the time being these PPPs are still largely input based rather than output based and therefore do not account for differences in productivity between countries. Furthermore, PPPs that include health care prices may be hard to interpret due to technical change in the sector and different regulatory regimes across countries.

contain further sub-groupings – or intermediate indicators – assessing the impact of six types of policies (Figure 6). As concerns quality in resource allocation, the two most important intermediate indicators are the degree of decentralisation and the extent of matching resources to specific needs. Under the heading of quality in budget management, two crucial efficiency-enhancing characteristics are the extent to which policy is outcome-focused and the degree of managerial autonomy especially at the school level. The quality in market framework is presumed to be related to the degree of competitive pressure in service provision which involves the presence of market signals, namely, benchmarking and schemes allowing for user choice.

Figure 5. **Estimate of cost efficiency at the national level**

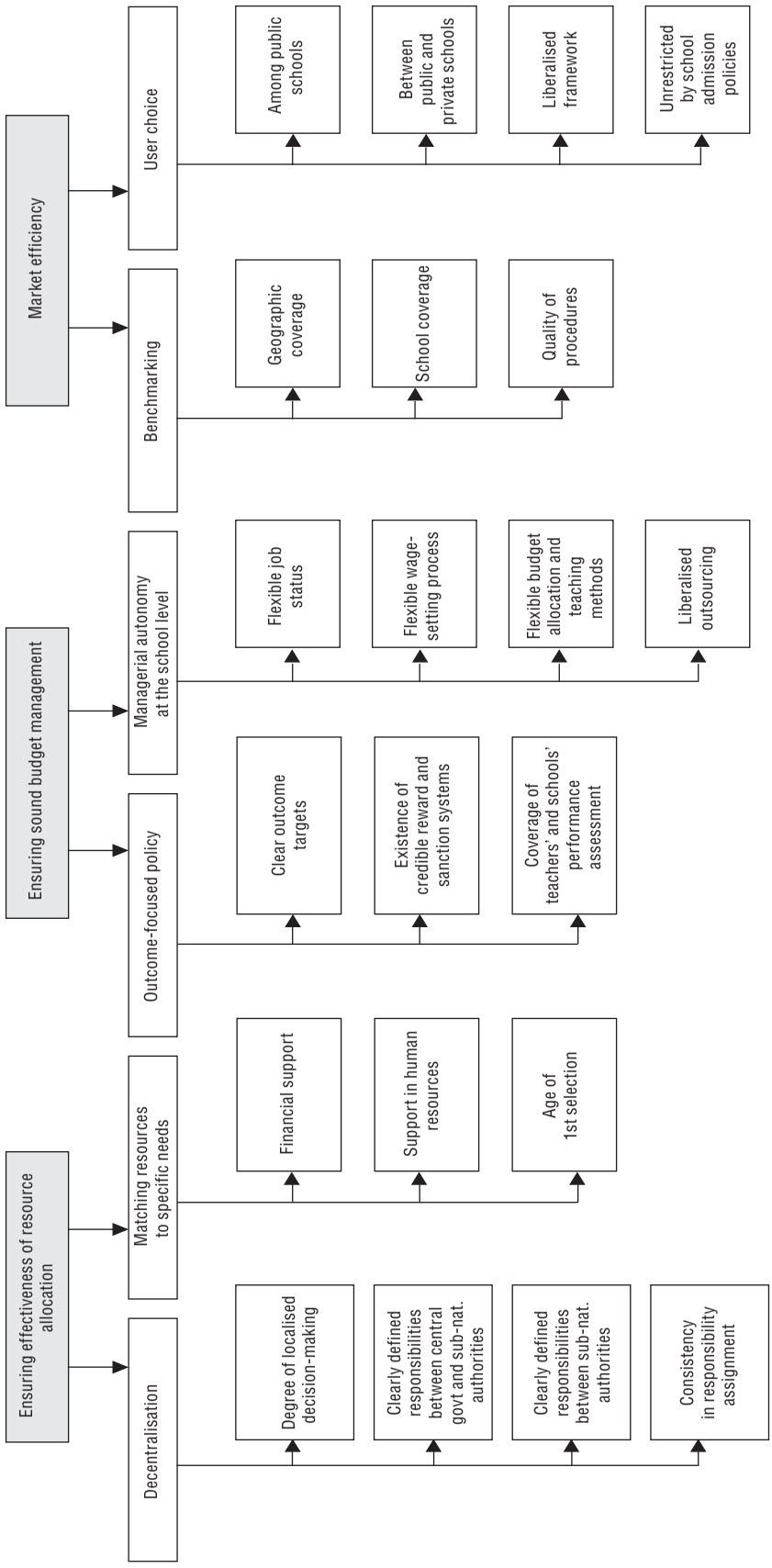


Note: DEA performed with 2 inputs (cumulative expenditure per pupils and pupils' socio-economic background) and 1 output (average PISA score). The figure shows the bias-corrected estimate of efficiency and 95 % confidence intervals.

1. Data for these countries concern public institutions only.

Based on questionnaire responses, 21 low-level indicators were constructed for 21 different policy domains. To compute the six intermediate indexes requires an assessment of the relative importance ascribed to the different low-level indicators, which may vary both across countries, given different institutional settings, and over time. To help overcome the difficulties in assigning relative importance to individual aspects of institutional and policy efficiency in widely different budgetary frameworks, the approach adopted here is based on random weights, which assumes complete uncertainty about the appropriate value of each of the individual weights used in the indicator construction. In total 1 000 indicators were calculated for each intermediate indicator, with the differences between the indicators being the (random) weights used in each construction. This

Figure 6. **Structure of institutional indicators**



technique also sheds light on the consistency of policies in that there will be greater variation in the composite indicators when individual policies are relatively inconsistent in a given policy domain.

Quality in resource allocation

The degree of decentralisation of responsibilities between central government and sub-national public authorities is assumed to enhance efficiency in the allocation of public spending resources insofar as educational needs may differ from one geographical area to another and resources should be matched to them. However, decentralisation may become counterproductive and reduce efficiency if it is poorly designed, resulting, for instance, in overlapping responsibilities between levels of government. In this context, the policy setting is assessed by the degree of localised decision-making power and the extent to which the allocation of this power is clearly defined and consistent.

Matching resources to specific needs – which can encompass mechanisms to support the disadvantaged – may have a favourable impact on overall educational efficiency, notably by avoiding "cream-skimming" effects at the aggregate level. Such mechanisms may be required in order to make up for the tendency of education systems to under-provide services for less able pupils. Thus, policies in this area concern whether specific educational needs are taken into consideration in financing schools and rewarding teachers as well as the age that tracking is implemented.

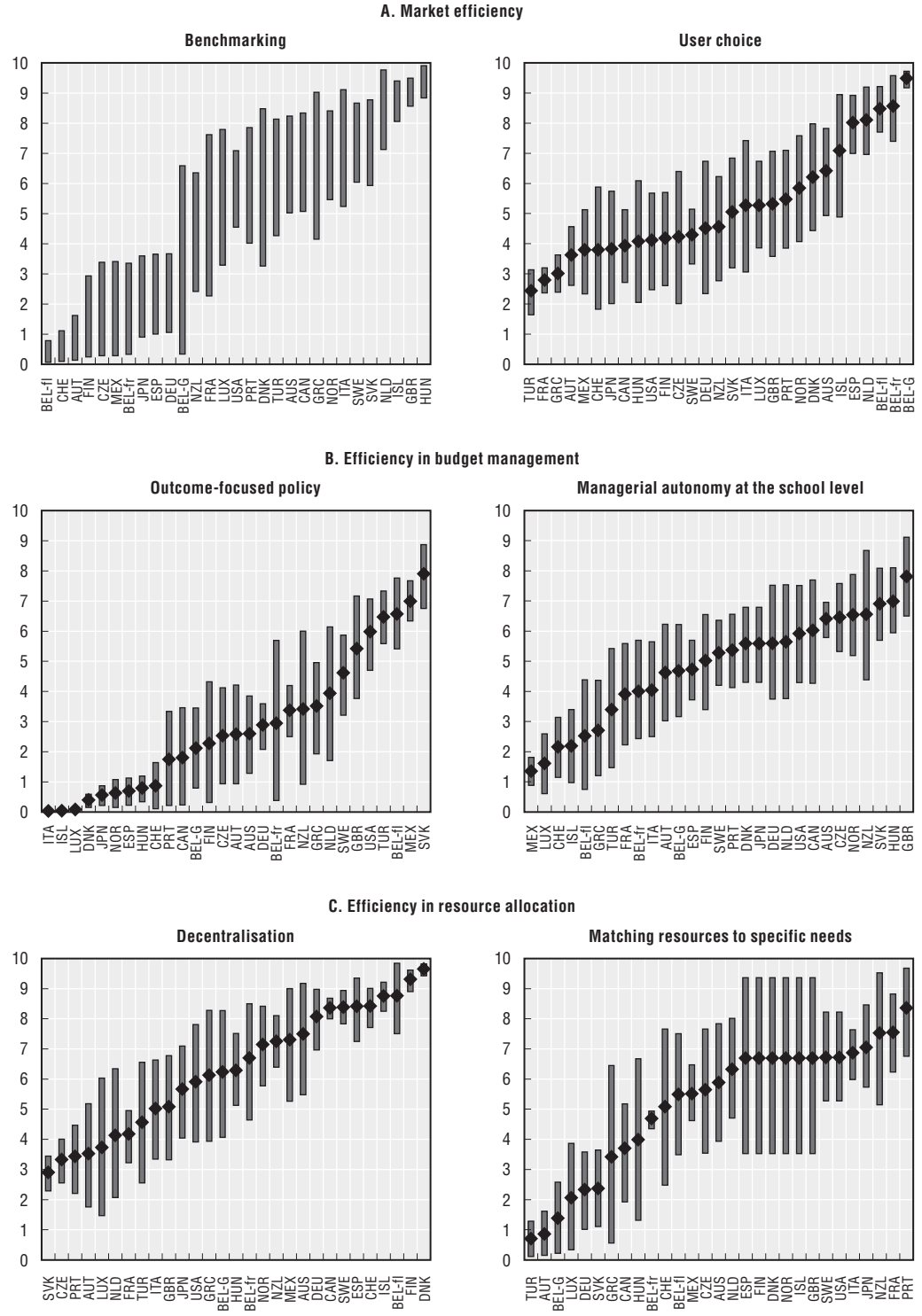
The intermediate indicators for decentralisation and matching resources to specific needs are presented in Figure 7 and show substantial differences in policy settings across countries. The figures show the mean value and the range that contains 90% of the 1 000 intermediate indicators. A small spread of the range that contains 90% of the intermediate indicators can be seen as a measure of the consistency of a country's policies in that domain. The indicator for decentralisation is high for Denmark and Finland, where policies are also relatively consistently favourable for decentralisation. At the other extreme, policy settings appear relatively less decentralised and consistently so in the Czech Republic, Portugal and the Slovak Republic. The ability to match resources to specific needs appears rather well developed in France, Portugal and New Zealand and relatively underdeveloped in Austria, Germany, Luxembourg, the Slovak Republic and Turkey. Comparing the variation of indicators reveals that policies appear less consistent in matching resources to specific needs in a number of countries.

Quality in budget management

The extent to which policy is outcome-focused allows clear objectives to be set for public institutions involved in education, especially if backed by associated evaluation, reward and/or sanction systems. Policies in this area include whether clear targets are established, whether there is a credible reward and sanction system in place, and the coverage of performance assessment mechanisms.

The degree of managerial autonomy, especially at the school level, based on flexibility of job status, wage setting and budget allocation and disciplined by liberalised outsourcing, where possible, may also make for greater efficiency in the use of resources. Policies considered in this area consist of the extent of flexibility in managing human resources, setting wages, allocating budgets and teaching methods as well as the opportunity to use outsourcing.

Figure 7. **Estimates of institutional indicators**



Note: The figure gives the average and the range that contains 90% of the 1 000 random weighted indices. Indicators for Belgium are presented for each of the communitites: German (BEL-G), Flemish (BEL-Fl.), French (BEL-Fr.).

Source: OECD Secretariat calculations based on questionnaire responses.

The intermediate indicators for outcome-focused policy and managerial autonomy are presented in Figure 7. Some countries, such as Iceland, Italy and Luxembourg, define few or no objectives, whereas countries such as the Slovak Republic and Mexico appear to favour strong outcome-focused management practices. With respect to managerial autonomy, Mexico along with Luxembourg, Switzerland and Iceland appear to grant public school managers a relatively limited degree of autonomy. By contrast, school managers in the United Kingdom, Hungary and the Slovak Republic appear to enjoy significantly more autonomy.

Quality in market framework

Competitive pressure can act as a spur to greater efficiency in service provision, which requires the presence of market signals. In this respect, *benchmarking* may improve service provision by identifying best practices and inefficiencies that can act to increase competitive pressures. Benchmarking is assessed against its coverage and quality in the composite indicators. While, benchmarking will have a positive impact, there may be an element of endogeneity in this measure if education authorities resort to benchmarking when performance is poor.

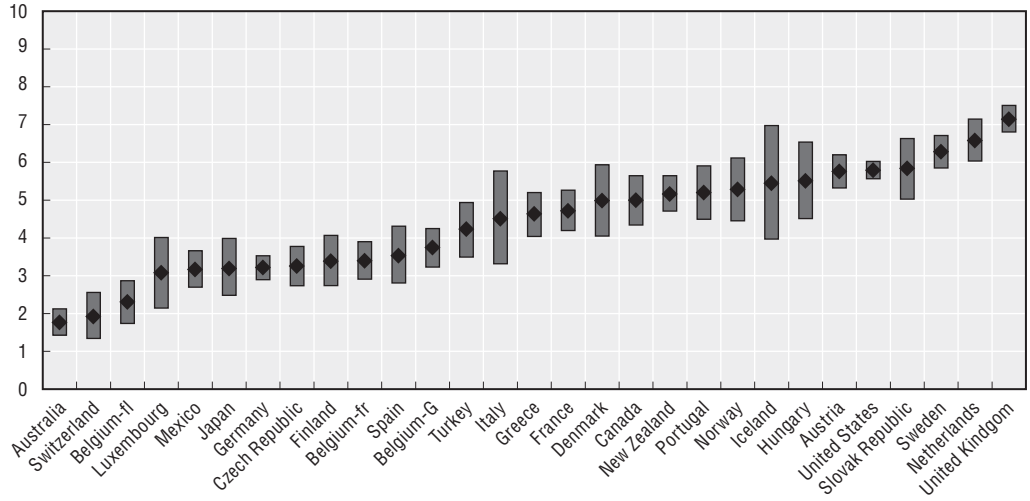
Allowing for user choice among alternative providers of educational services may be one of the most effective means of giving market signals a role in enhancing the effectiveness of public spending in education. This may strengthen competitive pressures and result in services which respond better to citizens' needs – provided that spending follows the user. In this light, policies in this sector include allowing choice among alternative schools and the extent to which this choice may be constrained.

The intermediate indicators for benchmarking show marked variation in the use of benchmarking across the OECD, ranging from almost none in Austria, Belgium (Flemish community) and Switzerland to extensive coverage and high quality indicators in Hungary, Iceland, the Netherlands and the United Kingdom. All countries appear to allow user choice, but the extent is relatively limited in France, Greece and Turkey. In Belgium, the Netherlands and Spain – three countries where private schools represent a sizeable share of schools – user choice is well developed.

Assessing the overall quality of the institutional environment

One approach to assessing the results is to aggregate the information into overall composite indicators, which also take into account possible complementarities between various policies. Four types of complementarities between the intermediate indexes have been taken into account in the composite indicators. There complementarities relate to outcome-focused policy with managerial autonomy at the school level and/or benchmarking; managerial autonomy with outcome-focused policy; user choice with the quality of benchmarking, and decentralisation with benchmarking.[9] The aggregate composite indicators (Figure 8) suggest that the most favourable institutional frameworks for enhancing efficiency in the education sector are in the Netherlands, Sweden and the United Kingdom. The institutional frameworks in Austria, Belgium (Flemish community) and Switzerland appear less conducive to promoting efficiency.

Figure 8. **Composite institutional indicator**



Note: The figure gives the average and the range that contains 90% of the 1 000 random weighted indices. Indicators for Belgium are presented for each of the communitites: German (G), Flemish (Fl.), French (Fr.).

Source: OECD Secretariat calculations based on questionnaire responses.

A more detailed examination of the various indicators in Gonand *et al.* (2007) suggests four possible country groupings:

- In a first group of countries, including the United Kingdom, Australia, Norway, Denmark and the Netherlands, the characteristics of the institutional framework seem to be very favourable, compared with OECD average, with respect to market efficiency (*i.e.* benchmarking and user choice) and to either resource allocation efficiency or budget management efficiency. For these countries, performance in any efficiency type is never lower than the OECD average.
- In another group of countries, which includes the United States, New Zealand, Iceland and Portugal, the institutional setting in primary and lower-secondary education remains globally favourable to spending efficiency, albeit national performance can be lower than the OECD average in one specific efficiency type, or higher in only one area out of the three considered in this paper.
- The overall picture is mixed for France, Slovakia, Italy, Canada, Belgium (Flemish community), Finland, Spain and Mexico, where the institutional characteristics of the public education sector either do not stand out as more (or less) favourable in any of the three dimensions of spending efficiency, or where outperformance in one category is offset by a lower score in another one. Germany, which was identified as relatively unfavourably placed by the composite indicator, belongs to this group.
- Results are less favourable for the Czech Republic, Greece, Luxembourg, Japan, Turkey, Hungary, Belgium (French speaking community), Switzerland and Austria. Among these countries, Austria, Switzerland, Japan, Luxembourg and the Czech Republic also ranked relatively low with respect to the composite index. This group includes several countries which are low in market efficiency (Austria, Switzerland, Japan and the Czech Republic), while efficiency of resource allocation is a potential problem in Austria, Greece, Luxembourg, Hungary and Turkey. The institutional framework is assessed as being

relatively unfavourable in respect of both market efficiency and budget management in Switzerland, and unfavourable in respect of all efficiency types in Austria.

The empirical evidence of policies promoting greater efficiency

The ultimate objective of the research undertaken here is to evaluate the possible influence of institutional settings on the measures of efficiency performance. To that end, the estimates of efficiency are used as the dependent variables and the measures of institutional quality as the explanatory variables in regressions. In addition, in order to exploit the much larger dataset of school level efficiency estimates, responses to the PISA 2003 school-level questionnaire can be used to test the influence of policies, which are components of the country-level institutional indicators. In particular, the PISA responses give school-level information with respect to aspects of: decision-making autonomy; student assessment and streaming and school selectivity, as well as types of ownership.[10]

The empirical approach adopted faces two main obstacles in identifying the influence of policy on performance. One drawback is that the indicators for the institutional environment are based on information concerning 2006, though the estimates of efficiency are based on education received mainly during the 1990s, as captured by the PISA 2003 survey. This may not be too problematic if the institutional framework of education systems or for individual schools has been relatively unchanged. But, in some cases, quite radical reforms have been implemented over the past decade. A second limitation is the cross-sectional nature of the data. Results are limited to identifying empirical relationships that lend plausible support to the possible role of institutional settings in affecting efficiency, but causality could be different.

The two-stage procedure of first assessing efficiency and then assessing the institutional and policy influences on efficiency was adopted because the non-parametric techniques used to estimate levels of efficiency are sensitive to misspecification and the accuracy of the estimation falls dramatically as the number of variables used in the specification rises. Inasmuch as the institutional variables were *expected* to influence efficiency, the two-step approach provides a framework to test for significance (Box 4 outlines the approach to do this using truncated regressions). While a two-step approach has limitations – there may be biases in either direction due to the omission of policy variables from the estimates of efficiency – results from parametric regressions using stochastic frontier analysis are also presented in the school-level analysis to check the robustness of the findings.[11]

Country-level institutional settings and efficiency

An examination of the relationship between institutional indicators and efficiency indicators at the country level is based on an extremely small sample of countries. In many cases, estimates failed to show any statistically significant relationship between the institutional indicators and the measures of efficiency, particularly for the aggregate indicators of institutional settings described in the previous section.[12] However, a set of regressions at the country level examines whether the performance indicators are related to the extent to which decisions are devolved to the school level, using information from *Education at Glance* (OECD, 2004b). (This component of decision making is included in the resource allocation indicator.) It should be noted that these regressions being based on information on the decentralisation of decision-making in 2003 to some extent overcome the problem of the discrepancy in timing of the measure of performance and the indicators

Box 4. **Estimation procedure using truncated regressions**

The distribution of efficiency estimates is truncated at the point of observed best practice and this needs to be taken into account in the estimation procedure. Maximum likelihood techniques are available to estimate truncated regressions.* Measurement error affecting observations on the efficiency frontier can create complicated patterns of serial correlation for observations that lie within that portion of the frontier. This is particularly a problem at the tails of the distributions that determine the efficiency frontier. One possibility to address this – adopted recently by a number of authors – is to bootstrap the truncated regression model (Simar and Wilson, 2007; Afonso and St. Aubyn, 2006). "Pseudo data" are generated from a truncated normal distribution and repeatedly used as the dependent variable using the same explanatory variables.The resulting estimates are then used to construct confidence intervals for the point estimates from the original truncated regression.

* Alternative approaches used to explain inefficiency, such as Tobit analysis, which is often used in the empirical literature, will lead to inconsistent estimation (Maddala, 1983). The approach adopted in this paper does not take account of the possible truncated distribution of the institutional indicators or the fact that the dummy variables are categorical (that is the variable is either 0 or 1). Zelenuyk (2005) reports that the test for significance for the coefficients of dummy variables will tend to show statistical significance only when the differences in the efficiency levels between the groups identified by the dummy variables are large.

of institutional settings mentioned above. These regressions reveal a reasonably consistent picture of better education performance in countries where decentralisation is more pronounced (Table 3).[13] The potentially beneficial effects of autonomy over the organisation of instruction (columns 2 and 7) and, to a slightly lesser extent, planning – including such decisions as whether to open or close a school and programme design – (column 4) are particularly striking.

The composite institutional indicators and intermediate indicators do, however, play a more positive role in explaining the *variation* in school-level efficiency scores within countries (Table 4). The results suggest that there is less variation in school-level efficiency

Table 3. **Cross-country correlations between efficiency scores and decentralisation**

	Technical efficiency for the median school									
	Input orientation					Output orientation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share of decisions made at school-level										
All domains	+++					..+				
Instruction		+++					+++			
Personnel			..+					..+		
Planning				+++					..+	
Resources					…					…
Relative teachers' remuneration	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++

Note: + signifies a positive correlation with higher levels of efficiency, a - signifies a negative correlation, while a dot indicates that the relationship was not statistically significant. The results are for efficiency measures using different assumptions about returns to scale. Thus, if a positive correlation holds for an efficiency indicator using all assumptions about returns to scale, the cell will record three plus signs.

when there is a more favourable institutional setting, as measured by the composite indictor. The indicator is negatively correlated with the measure of variation in efficiency performance in two of the three specifications reported (corresponding to the different methods of aggregation employed). Policies that score more highly in the domains of resource allocation are particularly strongly correlated with improved performance within a country.[14]

Table 4. **Within-country variation of efficiency and institutional settings**
Cross correlations between institutional indicators and the range containing the middle 90% of efficiency scores

	Cross correlations	
	Input orientation	Output orientation
Composite indicator	–0.340*	–0.314
Budget management	–0.184	–0.027
Market efficiency	–0.255	–0.255
Resource allocation	–0.340*	–0.446**

Note: A negative sign indicates lower levels of inefficiency. * and ** signify that the cross correlations are significant at the 10% and 5% level of significance, respectively. The institutional indicators were constructed using multiplicative aggregation in order to account for complementarities between policies.

School-level policy settings and efficiency

The results for efficiency performance at the school level are built on multivariate baseline truncated regressions (Table 5). The dependent variables are the input-saving and output-increasing estimates of school-level technical efficiency calculated from the PISA 2003 school-level database.[15] The explanatory variables correspond to the components of the institutional indicators of policy settings and are derived from the same database. These variables include the extent of school-level autonomy; the use of streaming within the school; the use of assessment, and the importance of selection for the school; as well as variables describing the size of school; ownership status and proxy measures of the quality of teaching inputs, together with a set of country dummy variables. Where possible, the specific types of policies at the school level are examined in more detail (for example, looking at different types of selection policies rather than aggregate indices concerning the extent of selection). The following paragraphs discuss the results in more detail and present supplementary estimations.

In order to provide a check on the robustness of these results, Table 6 reports estimates of stochastic frontier estimates of translog production functions.[16] The specification uses average PISA attainment as the dependent variable and uses as principal explanatory variables the teacher-student ratio, the availability of computers, the average socio-economic and language background of students in the school, which is augmented by the additional explanatory variables used in the truncated regressions. In this context, the estimates report whether the different policy variables are associated with higher (positive coefficient) or lower (negative coefficient) levels of attainment, other things being equal.

Size, ownership and quality of resources

The results for the regressions reported in Table 5 and Table 6 suggest that the size of school matters for efficiency. The results from the truncated regression suggest that this holds particularly for differences in terms of potential savings in inputs.[17] Typically schools

Table 5. **Determinants of technical efficiency at the school level**

Effect on inefficiency

	Coefficent estimate	Confidence intervals derived from pseudo distribution	
		Lower (2.5)	Upper (97.5)
Factors affecting inefficiency			
A. Input orientation			
Size (x 100)	–0.014**	–0.016	–0.013
Government-dependent private	–0.014	–0.034	0.005
Independent private	0.054**	0.026	0.083
Teacher qualifications	–0.040**	–0.069	–0.012
Autonomy resources	–0.002	–0.006	0.002
Autonomy curricular	0.003	–0.003	0.008
Some streaming	0.019**	0.060	0.033
Complete streaming	0.034**	0.021	0.047
Assessment: medium	–0.009	–0.021	0.005
Assessment: high	0.001	–0.015	0.017
Selectivity: considered	0.009	–0.003	0.022
Selectivity: priority	–0.018**	–0.035	0.000
Selectivity: pre-requisite	–0.037**	–0.052	–0.021
B. Output orientation			
Size (x 100)	–0.006**	–0.007	–0.005
Government-dependent private	–0.020**	–0.035	–0.005
Independent private	0.000	–0.023	0.021
Teacher qualifications	–0.027**	–0.048	–0.005
Autonomy resources	0.003**	0.000	0.006
Autonomy curricular	0.001	–0.003	0.006
Some streaming	0.020**	0.009	0.030
Complete streaming	0.031**	0.020	0.041
Assessment: medium	–0.001	–0.011	0.009
Assessment: high	0.008	–0.004	0.021
Selectivity: considered	0.001	–0.008	0.011
Selectivity: priority	–0.033**	–0.047	–0.020
Selectivity: pre-requisite	–0.064**	–0.076	–0.052

Note: a negative sign indicates lower levels of inefficiency.

** Indicates that the confidence interval does not encompass zero, giving an indication of the statistical significance of the estimate.

in less populated settlements are smaller, around one third of the size of schools in more densely populated areas. Hence, geographical factors in some cases will make it infeasible to shift the distribution of schools towards larger schools, but this may not represent a binding constraint in a number of countries.

The results suggest that government-dependent private schools (those privately-owned schools that rely on state funding for the majority of their income) tend to be relatively efficient in the output-maximising direction. Levels of technical efficiency measured in the output direction are higher on average in government-dependent private schools, other things being equal, suggesting that a higher ratio of government-dependent schools could help raise system-wide efficiency. The level of efficiency appears to be sometimes lower in independent private schools when technical efficiency is measured in the input orientation.[18]

Consistent with the findings at the country level, indicators for higher quality teaching resources appear to be correlated with better performance at the school level. The results

Table 6. **Efficiency and school-level policies: stochastic frontier analysis**

	Half-normal		Exponential	
	Coefficient	Cluster-robust Z score	Coefficient	Cluster-robust Z score
Constant	5.129***	49.34	5.087***	52.78
Teacher-student ratio	0.039	1.40	0.041	1.20
Computer availability	–0.014	–0.65	–0.011	–0.59
Socio-economic background	0.466***	14.06	0.468***	17.16
Language background	0.108***	4.93	0.109***	5.22
Teacher-student ratio*teacher-student ratio	–0.050***	–2.82	–0.047***	–2.70
Computer availability*computer availability	–0.009	–1.84	–0.008**	–2.08
Teacher-student ratio*computer availability	0.020**	2.13	0.018**	2.00
School size (x 100)	0.002**	2.47	0.002**	2.32
Government-dependent private	0.021***	2.59	0.022**	2.27
Independent private	0.001	0.07	0.004	0.23
Teacher qualifications	0.011	1.03	0.013	1.33
Autonomy resources	–0.006***	–2.67	–0.006***	–2.59
Autonomy curriculum	0.007	1.55	0.007	1.56
Some streaming	–0.004	–0.86	–0.003	–0.75
Complete streaming	–0.016**	–2.31	–0.015**	–2.80
Assessment: medium	–0.001	–0.38	–0.001	–0.38
Assessment: high	–0.012**	–2.08	–0.013**	–2.54
Selectivity: considered	0.000	0.02	0.000	0.03
Selectivity: priority	0.022***	2.82	0.021***	2.82
Selectivity: pre-requisite	0.046***	5.48	0.047***	5.04
Lambda	2.33		1.156	

Note: The table presents the estimated coefficients and associated cluster-robust Z-scores for a translog production function. The cluster-robust Z score takes into account the possible lack of independence of school policies within country. ** and *** signify that the estimated coefficient is significantly different from zero at the 5% and 1% level of significance, respectively. A positive coefficient signifies a positive relationship with average performance in PISA tests.

in Table 5 reveal that the proportion of the teaching staff certified with the national authorities is positively correlated with higher levels of efficiency in the truncated regressions, particularly in the input-saving direction.[19] The estimated coefficients in the stochastic frontier analysis, while also positively related to higher levels of efficiency, are statistically insignificant. Schools also tend to be more efficient when principals report that teaching resources, but not generally the physical infrastructure of the school, are in good order.

Quality in resource allocation: decentralisation, matching resources to needs

Contrary to the country results, where the devolution of decision-making autonomy to the school level (using the PISA school-level database) was associated with higher levels of efficiency, the measures of decision-making decentralisation derived from the PISA 2003 school-level database paint an inconclusive picture. Although the aggregate indicator for resource autonomy is significant in the output-raising orientation in the truncated regression and the stochastic frontier analysis, estimations for different types of decision-making autonomy were generally insignificant in the truncated regressions.[20] However, the stochastic frontier estimates suggest that student attainment is higher when schools have autonomy over determining teachers' salaries, but not when assessment policies are decentralised to the school level.[21] The analysis also suggests that schools implementing streaming demonstrate lower levels of technical efficiency in both the input and output

orientations, particularly if streaming is for all classes.[22] This suggests that there may be opportunities to raise system-wide efficiency by reassessing the extensive use of selection within schools.[23]

Quality in market framework: benchmarking and user choice

The results relating to the frequency of assessment suggest that it has little systematic relation to technical efficiency (Table 5). However, particular types of assessment can be related to efficiency performance. This process uses the same general framework as the baseline regressions, but substitutes the assessment variables by variables that are given a value of one when the school practises a particular type of assessment. The main relationships of these types of assessment to school-level technical efficiency are:

- Assessments made to inform parents appear to be positively correlated with higher levels of technical efficiency measured in the output orientation in the truncated regressions, but not statistically significant in the stochastic frontier estimates. The scope for boosting efficiency here is greatest in Turkey where only four-fifths of schools report using this type of assessment.
- Assessment of pupil progress through the school appears to be correlated with higher levels of efficiency in both the input and output orientations in the truncated regressions. The variable was also statistically significant in some stochastic frontier estimates, though not all.
- Assessment made to group students appears to be negatively correlated with technical efficiency in both the input and output orientation, which is consistent with the finding for streaming reported above. This variable was also statistically significant in some stochastic frontier estimates.
- Assessments may be used both for national comparison and for comparison with other schools, and thus allow for benchmarking.[24] While the results are not robust across all specifications, they suggest that both types of assessment are associated with higher levels of school-level technical efficiency, which is consistent with findings that benchmarking is associated with higher student attainment.[25]

Using an aggregate index of school selectivity already developed in the PISA 2003 database suggests that more demanding selection is associated with higher levels of efficiency (Tables 5 and 6).[26] When schools report making certain criteria a priority and, more especially, a pre-requisite for selection, levels of technical efficiency tend to be higher. This finding is subject to two caveats. First, while school selection may be good for schools individually, the aggregate effect in the country as a whole may be detrimental if the outcome is associated with negative externalities in other schools. Second, the type of selection is important as the individual school selection policies can either facilitate or block student or parental choice.[27] To explore this issue, the relationship between school selection policies and school-level technical efficiency was assessed by using the same general regression framework, but substituting the selection indicators by variables that are given a value of one when the school practises a particular type of selection. The main results for the different types of selection are.

- Selection based on residence, which limits student (or parental) choice, is negatively related to school-level technical efficiency, particularly in the output orientation. Residence-based selection criteria are reasonably widespread, with over half the schools

in Greece, Poland, Switzerland and the United States reporting that residence is a prerequisite for selection.

- School selection based on academic record, or to a lesser extent recommendations from feeder schools, is related to greater levels of efficiency (amounting to around a 6% improvement for a school with the median level of technical efficiency, other things being equal). This relationship holds if the sample is restricted to just public schools.

Summary of conclusions

The measurement of efficiency in education and the examination of how institutions can affect efficiency are hampered at the international level by a lack of comparable data. Choices need to be made that either limit the richness of the specification of the educational production function or limit the amount of information available over time. The fact that the results were comparatively weak at the national level may reflect both the limited number of policy indicators and the relatively few observations available. The small size of the national cross-section reduces the power of the non-parametric estimators, which are very data demanding. Notwithstanding these shortcomings, the research was able to identify important differences in efficiency across and within countries and also to link some of these differences to differences in policy settings across countries and also across schools.

The analysis demonstrates that a considerable variation exists in levels of educational efficiency across countries as well as across schools within (some) countries. At the school level, the results indicate that when controlling for the effects of socio-economic and linguistic background on educational attainment, resource savings would be as much as one third for the median school if the school raised its level of efficiency to that of the best performing school in the OECD. Even the more modest ambition of raising efficiency levels to the best performing *national* schools would give significant savings in resources. At the national level, controlling for the effects of socio-economic background on educational attainment, the results are broadly similar to the school-level results. Holding resources constant while moving towards best practice, PISA scores could be boosted by an average of 5% for OECD countries and by around 10% for the least efficient. The conclusions are similar for potential improvements in cost efficiency.

The analysis could not identify robustly aggregate institutional settings that were related to more efficiency in the primary and secondary education sector, but greater decentralisation appears to improve efficiency and policies which are conducive to better resource allocation reduce the variation in efficiency levels within countries. It also identified a number of policy settings at the school level that appear conducive to raising efficiency, such as greater decision-making autonomy and assessment policies that monitor student performance and allow for benchmarking between schools. Efficiency also tends to be greater when the proportion of teaching staff that are either certified or possess pedagogical qualifications is larger, while countries registering higher scores for the efficiency in budget management indicators tend to be more successful in restraining input use for a given level of educational attainment. Small school size, streaming and residence-based selection, which are widespread among OECD countries, tend to be associated with greater inefficiency.

The analysis suggests that for most OECD economies the scope for efficiency gains is proportionally greater from resource savings than increases in educational output. But this

is a comparative static construct and a complete evaluation of the policy options would need to consider the wider impact on economic growth, which this article does not do. Raising efficiency by boosting output and enhancing the human capital of the population can have beneficial impacts on labour force productivity, thus increasing GDP in the long run. On the other hand, resource savings might be used for cuts in taxes or switched to alternative investment uses, with consequences for economic growth which will differ with country circumstances.

Notes

1. Similar results are derived if alternative variables, such as measures of immigrant status, are used. Comparing the variables under stochastic frontier analysis estimation and using tests for variable relevancy in the DEA framework suggests that the language indicator is the preferred variable.
2. Other outputs used included the ratio of the 25th percentile to the 75th percentile as an indicator of the homogeneity of PISA scores The percentage of students attaining basic competences in mathematics was also used in the analysis, but exploratory results did not suggest its addition added to the findings when using the average measure of the PISA scores on their own.
3. The estimation includes quadratic and cross-product terms for teaching inputs and computer inputs (serving as a proxy variable for capital inputs), and also controlled for the socio-economic and language background of the pupil. The likelihood ratio test also supported this functional form over the Cobb-Douglas production function. The final specification was partly determined by problems of multicollinearity encountered in an initial more general specification. Multicollinearity is a frequent problem for translog estimation.
4. Results are reported using the convention that 1 is a fully efficient observation and that inefficient observations are less than 1. These are the Farrell measure of input orientated efficiency and the Shepard measures of output orientation efficiency.
5. The overall rank correlation between the estimates for the baseline and wider specification is high, at around 0.9. Separate estimates of school-level efficiency were also made for individual country samples. While the estimates of inefficiency were generally lower – a feature expected from using smaller samples – there was little change in the relative rankings.
6. While Turkey emerges as one of the most efficient countries, these results should be treated with caution. Turkey has very high drop-out rates for its students: only 54% of 15-year-olds are offered education compared with an OECD average of 95%.
7. If public schools only are taken into account, the average cross-country input and output oriented efficiency levels remain unchanged. But this hides some country differences, notably for Belgium (the Flemish community), where efficiency estimates deteriorate, and New Zealand, where efficiency estimates improve.
8. This work drew on past empirical work that identified institutional determinants of public sector efficiency (Joumard *et al*., 2004).
9. Two methods of computing composite indicators and scores for each category of efficiency have been developed which take account of such complementarities. See Gonand *et al*. (2007) for details.
10. Apart from the composite indicators already available in the PISA 2003 database, the variables derived are mainly dummy variables, which assign a value of one if the school implements a given policy and zero otherwise.
11. These estimates report cluster-robust standard errors (z statistics) in order to take into account the possible lack of independence of policies across public and private schools within a country.
12. The median school is used in the assessment of country-level institutional settings not only to check against possible aggregation biases (Hanushek, 1996) but also because the point estimates of efficiency from the country-level sample are subject to considerable uncertainty. This is an unavoidable drawback of DEA analysis on small samples.
13. See Bishop and Woessmann (2004) on the importance of school-level decision-making autonomy. Barankay and Lockwood (2006) using data from Swiss cantons show that greater autonomy is correlated with higher levels of student attainment.

14. The negative correlation for the indicator of resource allocation occurs irrespective of the method of aggregation employed in its construction, suggesting a quite robust relationship.

15. The results reported assume non-increasing returns to scale.

16. Stochastic frontier analysis was used in Sutherland *et al.* (2007) to check the robustness of data envelopment analysis estimates of efficiency.

17. This finding was robust to the different measures of technical efficiency and different specifications reported below. Larger schools are systematically more efficient in all the specifications.

18. Fuchs and Woessmann (2004) note higher levels of student attainment in government-dependent private schools. The coefficient for purely private schools is less robust, being particularly sensitive to the inclusion of Japanese private schools in the estimation. Estimations omitting Japanese schools from the sample show that independent private schools tend to be more efficient than public schools.

19. The proportion of staff that is certified with the national authorities is used, as a larger number of schools responded to this question. If the variable used in the estimation is replaced by the proportion of teachers who possess teaching qualifications the results are very similar. Research focusing on attainment has found limited support for the proposition that teaching qualifications – an observable proxy for teaching quality – contribute to higher levels of output.

20. Examination of the individual responses to questions concerning decision-making autonomy in particular domains, such as whether the school can choose its own textbooks, was also inconclusive.

21. The finding that centralised assessment is conducive to better student attainment is a result noted in other analyses (Fuchs and Woessmann, 2004).

22. The institutional indicators include information on the age of first selection, which is argued to discriminate against students sorted into the "low" stream or track if selection is too early. The PISA school-level database does not include information on the age at which streaming commences, but there is information on the extent to which it is used within schools. If, however, streaming is implemented by shifting pupils between schools (tracking) this will not be captured.

23. The sorting of students between schools (tracking) may have similar impacts in some cases. Shuetz *et al.* (2005) note *early* tracking heightens the influence of family background on attainment, suggesting that tracking has a negative effect on attainment in the "low track". Estimates of the impact of ability grouping on student attainment using PISA 2003 data, found that it could vary markedly across countries (Carey and Ernst, 2006), which may be a consequence of associated resource allocation (West and Woessmann (2003).

24. Studies on academic attainment frequently highlight the importance of centralised testing systems (Bishop and Woessmann, 2004).

25. The relationship is not statistically significant in all specifications, which could be related to features of how benchmarking is designed. If it distorts teaching incentives benchmarking could lead to "teaching to the test" at the expense of developing other areas of cognitive ability (Lazear, 2006; Jacob, 2007).

26. While for this aggregate indicator this is the opposite of the expected, this may reflect that competitive pressures may have a greater impact in some schools where selection is strong. Gibbons *et al.* (2006), using student-level data to investigate attainment in British primary schools, find that competitive pressures driven by greater parental choice have more impact in faith-based schools, which have more freedom in selecting students, than schools in the state system where choice appears to have little effect after accounting for location and pupil sorting. Heckman (2000) and Hoxby (2003) stress the fundamental role competition can play in raising efficiency.

27. There are other factors – such as design of how to implement choice and supporting institutional features – that can affect whether a greater degree of school choice is effective (Hoxby, 2003).

Bibliography

Afonso, A. and M. St. Aubyn (2006), "Cross-country Efficiency of Secondary Education Provision: A Semi-parametric Analysis with Non-discretionary Inputs", *Economic Modelling*, Vol. 23.

Barankay, I. and B. Lockwood (2006), "Decentralization and the Productive Efficiency of Government: Evidence from Swiss Cantons", *IZA Discussion Paper*, No. 2477.

Barrow, L. and C. Rouse (2005), "Causality, Causality, Causality; the View of Education Inputs and Outputs from Economics", *Federal Reserve Bank of Chicago*, *Working Paper*, No. 2005-15.

Bassanini, A., S. Scarpetta and P. Hemmings (2001), "Economic Growth: the Role of Policies and Institutions. Panel Data Evidence from OECD Countries", *OECD Economics Department Working Papers*, No. 283.

Bishop, J. and L. Woessmann (2004), "Institutional Effects in a Simple Model of Educational Production", *Education Economics*, Vol. 12.

Carey, D. and E. Ernst (2006), "Improving Education Achievement and Attainment in Luxembourg", *OECD Economics Department Working Papers*, No. 508.

Fuchs, T. and L. Woessmann (2004), "What Accounts for International Differences in Student Performance? A Re-examination using PISA Data", *Empirical Economics*.

Gibbons, S., S. Machin, and O. Silva (2006), "Choice, Competition and Pupil Achievement", *IZA Discussion Paper*, No. 2214.

Gonand, F. (2007), "The impact on Growth of Higher Efficiency of Public Spending on Schools", *OECD Economics Department Working Papers*, No. 547.

Gonand, F., I. Joumard and R. Price (2007), "Public Spending Efficiency in Primary and Secondary Education: Institutional Indicators", *OECD Economics Department Working Papers*, No. 543.

Hanushek, E. (1996), "School Resources and Student Performance" in G. Burtless (ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings Institution Press, Washington, DC, pp. 43-73.

Hanushek, E. (2003), "The Failure of Input-Based Schooling Polices", *The Economic Journal*, Vol. 113.

Hanushek, E. and J. Luque (2003), "Efficiency and Equity in Schools around the World", *Economics of Education Review*, Vol. 22.

Heckman, J. (2000), "Policies to Foster Human Capital", *Research in Economics*, Vol. 54.

Hoxby, C. (2003), "School Choice and School Competition: Evidence from the United States", *Swedish Economic Policy Paper*, Vol. 10.

Jacob, B. (2007), "Test-based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments", *NBER Working Paper*, No. 12817.

Joumard, I., P. Kongsrud, Y. Nam and R. Price (2004), "Enhancing the Effectiveness of Public Spending: Experience in OECD Countries", *OECD Economics Department Working Papers*, No. 380.

Lazear, E. (2006), "Speeding, Terrorism, and Teaching to the Test", *Quarterly Journal of Economics*, August.

Maddala, G. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge Univerity Press.

OECD (2004a), *Learning for Tomorrow's World: First Results from PISA 2003*, Paris.

OECD (2004b), *Education at a Glance*, Paris.

Shuetz, G., H. Ursprung and L. Woessmann (2005), "Education Policy and Equality of Opportunity", *mimeo*.

Simar, L. and P. Wilson (2007), "Estimation and Inference in Two-stage, Semi-parametric Models of Production Processes", *Journal of Econometrics*, Vol. 136.

Sutherland, D. and R. Price (2007), "Linkages between Performance and Institutions in the Primary and Secondary Education Sector", *OECD Economics Department Working Papers*, No. 558.

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

West, M. and L. Woessmann (2003), "Which School Systems Sort Weaker Students into Smaller Classes? International Evidence", *mimeo*.

Zelenuyk, V. (2005), "Power of Significance Test of Dummies in Simar-Wilson Two-Stage Efficiency Analysis Model", *Université Catholique de Louvain, Institut de Statistique Discussion Paper*, No. 0522.