Session Number: Parallel Session 8C
Time: Friday, August 27, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Estimating an Educational Production Function for Five Countries of Latin America on the Basis of the PISA data

Joseph Deutsch and Jacques Silber

For additional information please contact:

Name: Jacques Silber
Affiliation: Bar-Ilan University

Email Address: jsilber_2000@yahoo.com

Estimating an Educational Production Function for Five Countries of Latin America on the basis of the PISA data.

by

Joseph Deutsch*

and

Jacques Silber*

*** Department of Economics, Bar-Ilan University, 52900 Ramat-Gan, Israel.**
E-mail address of the authors: jdeutsch@mail.biu.ac.il;
jsilber_2000@yahoo.com

August 2010

Abstract

This paper takes a new look at the determinants of cognitive ability. Using the concept of education production function and on the basis of the results of the 2006 OECD PISA (Programme for International Student Assessment) survey for five Latin American countries (Braazil, Chile, Colombia, Mexico and Uruguay) it adopts an efficiency analysis perspective. This approach in itself is not new, but the present paper, rather than selecting as inputs a few variables among the many that are available in this survey, attempts to include the maximum amount of available information. Such an approach is made possible because in a first stage the huge amount of relevant information available on the production of education is aggregated via the use of Correspondence Analysis (CA). Before estimating the degree of efficiency of education production, an important distinction has to be made between what should be considered as discretionary inputs and which are really the only ones that should be taken into account when measuring the efficiency of educational production, and factors which are likely to have an impact on the efficiency of transforming these inputs into outputs. To measure efficiency we use the stochastic production frontier approach. However, rather than focusing on educational production by educational institutions, we analyze efficiency at the individual (student) level. Once such individual efficiency measures are obtained, we analyze via an OLS regression the determinants of this efficiency. In the final stage of the analysis, using the so-called Shapley decomposition, we attempt to determine the exact impact on efficiency of each of the non discretionary variables that are considered as determinants of efficiency.

Key Words: Brazil – Chile – Colombia - Correspondence analysis – Educational production function – Mexico - PISA data - Shapley value – Uruguay

J.E.L. Classification: A20 – O15

I. Introduction:

In most countries education is one of the main services provided by the government. The important role played by the state in the provision of education is certainly the consequence of the belief that education is an important source of human capital formation and hence of economic growth[1]. Several empirical studies have however shown some skepticism concerning the existence of a strong link between education and growth, specially in developing countries (see, for eample, Pritchett, 2001). It should nevertheless be stressed that many of the studies checking such a correlation between education and growth have been cross-country regressions where only the quantitative aspect of human capital (often measured as the average number of years of schooling) is taken into account. But when educational quality, as measured by international comparative tests of skills, is introduced, several researchers found quite a strong association of education with growth (see, Hanhushek and Kimko, 2000, and Barro, 2001). Note that, as far as countries from Latin America are concerned, the study of Behrman et al. (2008) suggests that both pre- and post-schooling experiences, not just schooling, have a significant impact on adult cognitive skills. If this is the case, schooling would be an imperfect representation of human capital in such cross-country growth relations.

Assuming that education, when correctly measured, turns out to have an impact on growth, it then becomes important to understand how it is produced. There have in fact been many studies of the educational production function, which often consider an educational institution as a firm transforming inputs into outputs. The inputs refer generally to the teaching and learning environment while the outputs are defined in terms of test scores. But here again some argue (see, for example, Pritchett and Filmer, 1999) that there is no strong evidence of an important impact of these inputs on educational outcomes. One of the problems faced by such studies is that the datasets on which they are based generally include only information on the contemporaneous family background and treat early childhood inputs as unobservables (for a very thorough review of this very important issue, see, Todd and

[1] This statement clearly assumes that education has positive externalities. We thank Jere Behrman for stressing this point to us.

Wolpin, 2003)[2]. The production of education is in fact a particularly complex issue in developing countries, as emphasized by Glewwe and Kremer (2006). They argue that whereas earlier studies found that education systems in developing countries had little impact on learning, more recent evidence based on natural experiments and randomized evaluations show a somehow different picture (see also, Glewwe et al., 2004, and Glewwe et al., 2007).

This paper aims at taking a new look at the determinants of cognitive ability. Its focus is on five Latin American countries (Brazil, Chile, Colombia, Mexico and Uruguay) that participated in the 2006 PISA survey. To estimate an educational production function it adopts an efficiency analysis perspective, an approach which is commonly used in productivity analysis and has been previously applied to the field of education (see, Worthington, 2001, for a survey of previous studies using such an approach). However rather than selecting as inputs a few variables among the many that have become available in recent surveys, such as the OECD PISA (Programme for International Student Assessment) survey, it attempts to include the maximum amount of available information. Such an approach is made possible because in this paper the huge amount of relevant information available on the production of education is aggregated via the use, not of Principal Component Analysis, a technique that should not be used with qualitative data, but of Multiple Correspondence Analysis. Before estimating the degree of efficiency of education production, an important distinction has however to be made between what should be considered as discretionary inputs and which are really the only ones that should be taken into account when measuring the efficiency of educational production, and factors which are likely to have an impact on the efficiency of transforming these inputs into outputs. To measure efficiency we use the stochastic production frontier approach but, rather than focusing on educational production by educational institutions such as schools[3], we analyze efficiency at the individual (student) level. Once such individual efficiency measures are obtained, we analyze via an OLS regression the determinants of this efficiency. In the final stage of the analysis, using the so-called Shapley decomposition, we attempt

[2] Some micro studies do however include early childhood factors (e.g. Behrman et al., 2008, and Todd and Wolpin, 2007, both of which found these factors to be important).

[3] See section II below for a survey of studies that adopted this approach.

to determine the exact impact on efficiency of each of the non discretionary variables that are considered as determinants of efficiency.

This paper is organized as follows. Section II gives a short survey of the applications of frontier efficiency measurement to education. Section III describes the methodology used in the present study. Section IV finally presents the three stages of the empirical investigation. We first limited the analysis to Colombia which was the only Latin American country for which all the variables defined in the 2006 PISA survey were available. Then we made a separate analysis for each of the five Latin American (Brazil, Chile, Colombia, Mexico and Uruguay) countries on the basis of the variables that were available in all five countries. Finally we estimated a regression common to all the countries, including dummy variables for the countries and on the basis of the results of this regression we implemented a Shapley decomposition procedure. The paper has five Appendices. Appendix 1 gives the list of the PISA questions that were used to determine the discretionary as well as the non discrtionary inputs. Appendix 2 explains shortly what correspondence analysis (CA) is. Appendix 3 summarizes the main elements of efficiency analysis and of its implementation via stochastic production frontiers. Appendix 4 describes what is known as the Shapley decomposition procedure. Appendix 5 gives the results, for each country, of the regressions estimating the determinants of individual efficiency. The "factor loadings"[4] for the aggregate inputs (whether they are discretionary or not), that were derived from correspondence analysis, are available from the authors upon request.

II) Applications of Frontier Efficiency Measurement to Education: A Short Survey of the Literature

"Education is a complex process with multiple objectives and inputs from school and home and also from personal characteristics of the students" (Waldo, 2006). Most applications of efficiency analysis to education have considered the educational institutions as firms transforming inputs into outputs, the inputs being the

[4] In the literature on correspondence analysis "factor loadings" are called dimensions. These "factor loadings" allow one to determine the coordinates of the observations and variables in a bi-plot based on the two first "factors" and to detect the similarity between the variables, between the observations or between the variables and the observations.

characteristics of the teaching environment and the outputs the students' test scores (for an earlier survey of the topic, see, Hanushek, 1986). The stochastic production frontier approach (e.g. Deller and Rudnicki, 1993, and Bates, 1997) includes a disturbance assumed to represent noise and measurement error and thus allows decomposing the deviations from the efficiency frontier into two components representing respectively inefficiency and noise. On the other hand the so-called programming approach (to which Data Envelopment Analysis, the "Free Disposal Hull" and the "Directional Distance" approaches belong) is not stochastic and as a consequence the deviation from the efficiency frontier is assumed to be entirely the result of inefficiency. The main advantages of this programming approach is that it can include multiple outputs, does not necessitate that one selects relative weights of the inputs and outputs and does not require specifying a parametric functional form for the production function.

Quite a few studies devoted to educational efficiency have used the stochastic frontier approach (see, for example, the analysis by Dolton et al., 2001, of the effective use of student time in producing educational performance) but there are also many studies of educational efficiency based on the programming approach (see, Worthington, 2001). Very often, when efficiency analysis is applied to education, the inputs are assumed to be the number of persons in the teaching or administrative staff as well as the educational expenditures, at the exclusion of the labor costs (e.g. physical facilities, library holdings, computers available, …). Outputs can be of a quantitative nature (number of graduating students, the percentage of students with specific attainments) but they may also be the consequence of the quantitative as well as of the qualitative aspects of schooling. This is the case of test scores in reading mathematics, writing, spelling (see, however Gstach et al., 2003, who argue in favor of a distinction between quantity and quality). The units to which efficiency analysis has been applied have been primary and secondary schools (e.g. Bessent et al., 1982, and Deller and Rudnicki, 1993), universities (Athanassopoulos and Shale, 1997) or university departments (e.g. Johnes and Johnes, 1993 and 1995, Madden et al., 1997, Johnes, 2006). Some of these studies assumed constant returns to scale while others supposed that there were variable returns to scale.

Evaluating the efficiency of educational units such as schools is however not an easy task. Portela and Camanho (2007) thus emphasize that "schools provide a service that has the usual characteristics of services like intangibility and heterogeneity, which hamper standardization, and the educational service is carried out on the actual pupil, who is at the same time an input and an output of the production process"[5]. Mancebón and Bandrés (1999) stress other difficulties, such as the time dimension of the educational process (the impact of education is usually revealed years later) and its cumulative nature (students' achievements are also a function of what happens in schools previously attended by the student) as well as the central importance of exogenous factors such as the student's family socioeconomic background and his/her ability.

Once the degree of (in)efficiency of each decision making unit has been determined the task of the researcher is to discover the determinants of such inefficiencies. These generally include inputs beyond institutional control such as students' talents and socioeconomic status. Other determinants may be the percentage of minority students, of students coming from single-parent households, the degree of competition between decision making units (see, for example, Duncombe et al., 1997, Grosskopf et al., 2001, or Waldo, 2006), the degree of administrative pressure or the ability of citizen to monitor costs (see, Worthington, 2001, for more details). The importance in an educational production function framework of making a distinction between inputs that can be varied by the decision making unit and those that are not discretionary was first stressed by Ruggiero (1996). In fact Portela and Camanho (2007) argue that school performance may be evaluated from two different perspectives. The first one which they call the society perspective, amounts to asking which school is the best one, assuming the parents's goal is the academic achievement development of their children. Schools are therefore supposed to promote the students' achievements, *given the socioeconomic background and the abilities of these students*. The alternative is to take an educational authorities' perspective, in which case schools should promote the students' achievements, *given the socioeconomic background of the students, their abilities and the school resources*. Clearly in the latter case "schools with less resources are required less in terms of achievement than school with more resources" (Portela and Camanho, 2007).

[5] It may not necessarilty be helpful to characterize students as outputs of the educational process. The real output is the student's knowledge and this is not the same as the student's time, for example, which is an input. We thank Jere Behrman for stressing this point.

Whereas many authors have analyzed the efficiency of the educational process by focussing their attention on schools, universities or university departments as decision making units, there have also been studies aiming at making international comparisons of the educational system as a whole.

Wilson (2005) based his analysis on the 2000 PISA database. Three types of inputs were used. The first one was assumed to measure the socio-economic status of the student and included two variables, an index of home education resources (derived from the students' reports on the availability in their home of a dictionary, a quiet place to study, a desk for study, textbooks and the number of calculators in their home) and an index of family wealth (derived from the students' reports on the availability in their home of a dishwasher, a room of their own, educational software, a link to the internet, and the number of cellular phones, television sets, computers, motor cars and bathrooms at home). Both indices were observed directly in the student file (Wilson, 2005) and averages of these individual indices were estimated at the school level. The second type of inputs was called aggregate physical inputs and was calculated as being equal to the sum of full-time teachers plus one-half times the number of part-time teachers. The third type of input was called aggregate teacher quality and was computed as the proportion of certified teachers. The two last categories of inputs were obtained from the school files. Three output categories were distinguished. The first one included individual test scores (in mathematics, reading, reading-retrieving, reading-interpreting, reading-reflecting and problem solving in science). The second one was the individual grade level attained at age 15. These two individual output categories were averaged at the school level (see, Ruggiero, 2006, on measurement errors at the individual level and the smoothing effect of aggregation). The third output was the number of students in each school. In order to minimize the number of inputs and outputs the author made a principal component analysis of the six individual test scores and used the first principal component as the first output variable. Similarly a principal component analysis was applied to the two variables measuring the socioeconomic status of the individual and the first principal component was considered as measuring the socioeconomic status of the student. Wilson (2005) used both Data Envelopment Analysis (DEA) and the "Free Disposal

Hull" (FDH) approach to compute an efficiency score for each school in each country. He then presented tables giving, for each approach, the median and mean values of these scores in each of the 40 countries of the database.

A relatively similar study was conducted by Afonso and St. Aubyn (2005) who considered that educational achievement, the output, is measured by the performance of 15-years olds on the 2003 PISA reading, mathematics, problem solving and science literacy scales. They selected as inputs the total intended instruction time in public institutions in hours for the 12 to 14-years old, averaged for the period 2000-2002, and the number of teachers per student in public and private institutions for secondary education, the calculations being based on full-time equivalents and averaged for 2000-2002. Their study covered 25 countries. To explain the efficiency scores obtained by each country they regressed this score on the 2003 GDP per capita and on parents' educational attainment assumed to be equal to the percentage of the population aged 35 to 44 that attained at least upper secondary education in 2001-2002. These two explanatory variables appear to be highly and significantly correlated with output scores.

The study by Waldo (2007) is exceptional in the sense that it applies efficiency analysis not at the level of the school but at the individual level. The author argues that such an approach makes it then possible "to carry out studies on a wide range of student constellations; e.g. entire educational programmes, schools or smaller groups within a school or a programme". The second originality of this study is that it uses directional distance functions so that efficiency is estimated as the distance of an observation to a production front defined by the best performing observations in the sample. The study uses an output-oriented model, the argument of the author being that students face given inputs offered by the school when determining their effort level. They certainly cannot reach efficiency by contracting inputs. The author uses Swedish data. Two outputs were introduced: grade value in core courses and grade value in profile courses[6] (the score is the sum of all grade values for the different

[6] Each program contains a number of courses attended by the students. The course credits given for a course depends on its difficulty relative to other courses. Each program has three types of courses. All students do a number of core courses common for all programs. These are Mathematics, Swedish, English, General Science, Social Studies, Religion, Art, and Physical Education and Health. The educational orientation of a program is set by a number of *profil*e courses that are compulsory for all students attending the program. Each program also leaves a number of credits to individual choice

courses attended by the student). The inputs selected were the student's grade in compulsory school, the mean number of full time equivalent teachers per student over the three years of upper secondary education, the mean grades from compulsory education of all students attending the same program at the same school as the student under evaluation (the idea is to measure the influence of peers) and a variable measuring the educational background of the parents. The efficiency of various programs is then estimated separately for both genders. It should be stressed that In the estimation of efficiency a student is never compared with students who have a higher input in any dimension. Thus, a student is never regarded as inefficient because his/her performance is lower than that of a student with higher compulsory school grades, better peers at school, or with more teaching resources.

The present paper, like Waldo's (2007) paper, measures efficiency at the individual level. However, rather than selecting a few inputs, it integrates, via correspondence analysis, a maximal amount of the information available in the PISA surveys. Moreover it uses a stochastic frontier rather than a programming approach so that it enables to break down the deviations from the efficiency frontier into two components reflecting respectively inefficiency and noise. Finally, by applying the concept of Shapley decomposition, it allows one to determine the respective contribution of the various explanatory variables to the dispersion of efficiency scores.

III) The Methodological Approach:

1) The first stage of the analysis: Data sources and data reduction procedure for the inputs:

A) Data sources:

The estimation of educational production functions in Brazil, Chile, Colombia, Mexico and Uruguay is based on data collected in the 2006 PISA survey. The Program for International Student Assessment (PISA) is a system of international assessments that focus on 15-year-olds' capabilities in reading literacy, mathematics literacy, and science literacy. PISA[7] emphasizes functional skills that students have acquired as they near the end of mandatory schooling. It began in 2000 and is administered every three years. Each administration includes assessments of all three subjects, but assesses one of the subjects in depth. The most recent administration was in 2006 and focused on science literacy.

B) Deciding which PISA variables should be considered as inputs

The PISA surveys include mainly three types of questionnaires (besides the tests themselves) that are respectively filled by the school administration, the parents of the student and the student himself/herself. Whereas it is not difficult to decide what the outputs will be in the efficiency analysis to be conducted since four test scores are generally available for each student in the PISA survey (scores in reading, mathematics, science and problem solving), it is much less simple to select the inputs. At this stage, before any aggregation procedure takes place, the idea is to choose inputs that can be considered as discretionary. It turns out that potential information on these inputs can be found in the school, parents and students' questionnaires. Appendix 1 lists the variables that were available for this efficiency analysis. A first set of these variables was considered as inputs in the production process. The second set of variables was included in the second phase of the analysis when an attempt will be made to determine the factors that have an impact on efficiency.

We assumed that the inputs to be included in the analysis should have the feature that they can be considered as discretionary inputs. Three categories of inputs have been selected (see, Appendix 1):

- the educational means available at home
- the inputs of the school, which may be divided into two categories, the pedagogical characteristics of school and the physical and human capital available at school[8]

[7] The PISA surveys are launched, every few years, by the Organization for Economic Cooperation and Development (OECD).

[8] It could be argued that the school inputs are in a way inputs chosen by the parents when selecting the school attended by their child.

- the time inputs of the student, which can be classified into two categories, time devoted to informal learning or to formal learning:

Appendix 1 lists all the relevant variables in each of these categories. These variables have been aggregated using Correspondence Analysis (CA).

C) On Correspondence Analysis:

Correspondence analysis was introduced by Benzécri (1972) and his French school. It is an exploratory data analytic technique aiming at analyzing simple two-way (or multi-way) tables where some measure of correspondence is assumed to exist between the rows and columns. Correspondence analysis is extremely useful to transform a set of complex data into quite a simple description of almost all the implicit information provided by the data.

A very useful characteristic of correspondence analysis is that it allows one to obtain a graphical display of row and column points in biplots, which helps discovering some structural relationships that may exist between the variables and the observations[9].

Although correspondence analysis (CA) may be defined as a special case of principal components analysis (PCA) of the rows and columns of a table, one should stress that CA and PCA have each specific uses. Principal components analysis is a useful tool when one has tables consisting of continuous measurement, whereas correspondence analysis is typically applied to the case of contingency tables.

While the Chi-square test is the usual procedure adopted for analyzing, in a cross-tabulation, the degree of association between rows and columns, this test does not allow us to find out which are the important individual associations between a specific pair of row and column. Correspondence analysis on the contrary indicates how the variables are related and not simply whether there is such a link.

Assume a contingency table that has I rows and J columns. The plot given by a correspondence analysis gives then a set of (I+J) points, I points corresponding to the rows and J points to the columns. If two row points are close one can then conclude

[9] See Appendix 2 for more details on this technique.

that their conditional distributions across the columns are similar. Given the symmetry of the role played by lines and columns in correspondence analysis we can also conclude that if two column points are close on the bi-plot provided by the correspondence analysis it implies that their conditional distributions across the rows are similar. Like principal components analysis, correspondence analysis provides the researcher with principal components which are orthogonal. More specifically each component is a linear combination of the variables on one hand, the observations on the other. The coefficients of these variables (observations) for the first two components give us in fact the coordinates that allow us to plot these variables (observations) in the graph previously mentioned (for more details, see Appendix 2 and Johnson and Wichern, 1999, chapter 12). We limited ourselves to the first factor.

2) The Second Stage of the Analysis: Using the Stochastic Production Frontier Approach to Determine the Efficiency of each Student

On the basis of the five inputs previously mentioned and of the four outputs provided by the PISA survey (scores in reading, mathematics, science and problem solving) an efficiency analysis was implemented and an efficiency score attributed to each student. The four test scores were considered as the outcomes of a latent variable reflecting the cognitive ability of the student (for a similar point of view see, Heckman et al., 2006, and Urzua, 2007). This latent variable is evidently not observed and to implement a stochastic production frontier analysis we used a technique originally proposed by Lovell et al. (1994) and later adopted by Deutsch and Silber (1999), Deutsch et al. (2003) and Ramos and Silber (2005).

Let $x = (x_1,..., x_5)$ denote the vector of the five aggregated inputs derived from "correspondence analysis". Similarly let $y = (y_1,..., y_4)$ refer to the four educational achievement scores obtained by the individual on the four tests (reading ability, mathematical ability, science, problem solving ability). Lovell et al.'s approach (see, Appendix 3 for more details) amounts to estimating a translog output distance function expressed as

$$\ln(1/y_M) = a_0 + \sum_{j=1}^{5} a_j \ln x_j + (1/2)\sum_{j=1}^{5}\sum_{k=1}^{5} a_{jk} \ln x_j \ln x_k$$
$$+ \sum_{i=1}^{3} b_i \ln y_i + (1/2)\sum_{i=1}^{3}\sum_{h=1}^{3} b_{ih} \ln y_i \ln y_h + \sum_{i=1}^{3}\sum_{j=1}^{5} c_{ij} \ln y_i \ln x_j + \varepsilon$$

where the subindex M refers to one of the four test scores[10] (see, Lovell et al., 1994, for more details on the procedure)

Note that the value of the five inputs, derived from correspondence analysis, were negative for some of the individuals. In order to be able to use a translog production function we transformed these inputs as follows

$$x'_{jk} = \frac{[x_j - Min\{x_{jk},...,x_{jK}\}]}{[Max\{x_{jk},...,x_{jK}\} - Min\{x_{jk},...,x_{jK}\}]}$$

where x_{jk} is the value of input $j\,(j = 1$ to $5)$ for individual $k\,(k = 1$ to $K)$ and x'_{jk} is the value of the "transformed input".

The technique of COLS (corrected least squares) is then used to obtain estimates of the various coefficients (see, Appendix 3 for more details on the COLS technique). The modified residuals which are then derived provide output distance functions for each individual by means of the transformation

$$d(k) = e^{[(\max imum\ negative\ residual) - (residual\ for\ individual\ k)]}$$

This distance will by definition be smaller than one (since its logarithm will be negative or at most equal to zero) so that all individual input and output vectors lie on or beneath the frontier.

These output distance functions measure the efficiency with which individuals convert their inputs into "educational achievements". Since the maximum observed output distance function is unity by construction, the individual distance divided by the maximum output distance may be considered as a kind of relative productivity index (it is called the Malmquist Productivity Index in the literature; for more details, see, Deutsch et al., 2003).

3) Third Stage of the Analysis: the Determinants of Individual Efficiency

A) Deciding which PISA variables should be considered as determinants of individual efficiency and aggregating these variables, using Correspondence Analysis (CA)

In Appendix 1 we have listed and classified all the variables that seem to be the relevant determinants of individual efficiency. Note that, except for Colombia, information from the questionnaire filled by the parents as well as on the Information and Communication Technology (ICT) available was not provided by the surveys, so that when analyzing the PISA data of Brazil, Chile, Mexico and Uruguay, as well as when estimating efficiency in a unique regression with dummy variables for the countries, we had to limit ourselves to the following categories of variables[11]:

- the gender of the student
- the human capital of the parents[12]
- the material wealth of parents
- information on school governance, the latter covering several domains (school funding, degree of autonomy of school, the degree of transparency of information, the degree of homogeneity of the school and its location)[13]
- the importance of learning efforts in the eyes of the student[14]
- the self-rated ability of the student[15]

Using again Correspondence Analysis we derived aggregated values for most of these determinants of individual efficiency[16]. Then a regression was estimated to understand which of these determinants play an important role.

B) Analyzing the determinants of individual efficiency

---

[10] See, Lovell et al. (1994) for more details on this technique which allows estimating the efficiency of transforming inputs into outputs.
[11] In Appendix 1 we put an asterisk (*) before the questions that were available only for Colombia.
[12] As indicated in Appendix 1, the human capital of the parents is related to the age of the father and of the mother, their main job, the highest level of schooling they completed, their country of birth and the language spoken at home (when compared to that in which the test is conducted).

[13] Correspondence Analysis was implemented separately on each of the domains of school governance.
[14] The student is asked to say how important it is for him/her to do well in science, mathematics and the test language.
[15] When working only with the PISA data for Colombia, the following additional categories of variables were available: the scientific background of the child when he was 10 years old, his/her familiarity with ICT and the importance of science at home.

Calling ε_i the individual efficiency score obtained when using the stochastic production frontier approach, we can then estimate an OLS regression $\varepsilon_i = z_i\beta + u_i$ where z_i refers to the determinants of individual efficiency previously mentioned and u_i is the error term.

C) Estimating the relative importance of the various determinants of individual efficiency

At this stage we apply a Shapley decomposition procedure that allows us finding out which determinants contribute most to the variance of the individual efficiencies. Appendix 4 summarizes the main idea of the Shapley decomposition originally proposed by Chantreuil and Trannoy (1999), Shorrocks (1999) and Sastre and Trannoy (2002). The application of this procedure to the R-square of a regression has been suggested by Israeli (2007). It has been applied to a logit regression by Deutsch and Silber (2006) and D'Ambrosio et al. (forthcoming). The general idea is to find out what is the contribution of the various explanatory variables to the "explained" variance of the efficiency score when the latter is regressed on these explanatory variables. In a certain way the Shapley decomposition procedure can be considered here as measuring the contribution of a given variable by implementing a stepwise regression procedure where all possible orders of introduction of the explanatory variables are taken into account, including the cases where some of them are not introduced as explanatory variables.

It should be stressed that the approach selected emphasized the concept of the efficiency of transforming inputs (assumed to be discretionary) into outputs (the four categories of test scores). The Shapley decomposition was then applied to the determinants (non discretionary inputs) of this efficiency and therefore does not allow determining the respective impacts of the various inputs and of the various determinants of efficiency on the outputs (test scores) of the students. We are only able to determine the *impact of the non discretionary inputs on the efficiency of*

[16] For the gender of the student and the four dummy variables describing the location of the school we evidently did not use correspondence analysis.

transforming discretionary inputs into test scores. If determining these relative impacts is considered to be a crucial issue, it is possible to tackle it by estimating output distance functions where both the discretionary inputs selected and the determinants of efficiency (non discretionary inputs) would be considered as inputs. The last phase (the efficiency regression stage) of the previous analysis would then be dropped and the Shapley decomposition would attempt to determine the relative contribution of all the inputs appearing in this extended list of inputs on, say, some measure of the dispersion of the output distance functions[17].

IV) Results of the Empirical Investigation:

1) The Case of Colombia:

As mentioned previously, Colombia is the only Latin American country for which data on all the inputs (discretionary and non discretionary) were available. The efficiency analysis involved five inputs and hence many interaction terms since we used a translog function. We do not present the results of the estimation of the efficiency analysis[18]. We directly give the results of the OLS regression where the dependent variable is the efficiency value for each individual. The explanatory variables are the 16 determinants of this efficiency, eleven of which have been obtained via correspondence analysis while the five others are the gender of the student and the four dummy variables describing the location of the school[19]. Table 1 gives the results of this analysis of the determinants of efficiency for Colombia. We then reestimated the regression by dropping some of the determinants whose coefficients were far from being significant. The results of the second regression are given in Table 2.

In order to better understand which of these variables has an important impact we implemented a Shapley decomposition to the results of Table 2. As mentioned

[17] We did not implement such a procedure but may try to do so in future research.

[18] They may be obtained from the authors upon request .

[19] As indicated in Appendix 1 the location of the school refers to the community in which the school is located, the possibilities being a rural area, a small town (from 3,000 to 15,000 people), a medium size town (15,000 to 100,000 people), a city (100,000 to 1000,000 people) and a large city (more than a million).

previously the Shapley decomposition procedure amounts to computing the marginal contribution of each of the explanatory variables appearing in Table 2 to the R-square of the regression (see, Appendix 4 for more details on the concept of "marginal contribution" and more generally on the Shapley decomposition procedure). The results of this investigation are given in Table 3. It appears that the most important determinants are the location of the school which contributes 19.5% to the R-square[20] and the human capital of the parents which contributes 19.2% to the R-square[21]. Three other variables play an important role: the gender of the student (contributes 16.3% to the R-square), the importance of science at home, as evaluated by the parents (contributes 13.2% to the R-square) and the the self-rated ability of the student (contributes 10.6% to the R-square). Note also that the degree of autonomy of the school, one of the elements of what we called information on school governance, contributes 8.6% to the R-square while the scientific background of the child when he was 10 years contributes 5% to the R-square. Finally note that the degree of transparency of the information about the school and the degree of homogeneity of the school contribute respectively 3.0% and 4.8% to the R-square. Both variables describe also what we called school governnance.

The role played by these variables may be summarized as follows. We may first stress the impact of genetic factors which include two factors, the gender of the student and his/her ability (although it is a self-rated ability). Together these two determinants contribute 26.9% to the R-square. Then there is the role played by parents which includes three components, their human capital, the importance of science at home and the scientific background of the student[22] . Together these factors contribute

[20] Note that in the Shapley procedure we implemented we grouped all the dummy variables measuring the location of the school. In other words when this determinant had to be eliminated in the Shapley decomposition procedure (see Appendix 4) we eliminated together all the dummy variables measuring the location of the school.

[21] Note that our approach does not allow us making a distinction between the effects of the education of the parents and those of intergenerationally-correlated endowments which some recent studies consider as underlying much of the association between parental and child schooling (see, Behrman and Rosenzweig, 2002 and 2005; Black, Devereux and Salvanes, 2005; Plug, 2004; Plug and Vijverberg, 2003).

[22] The parents are asked to "thinking back to when their child was about 10 years old" and have then to say how often their child watched TV programmes about science, read books on scientific discoveries, watched, read or listened to science fiction, visited websites about science topics or attended a science club. This is why the variable describing the scientific background of the student was classified in the category "role played by the parents".

37.4% to the R-square. Finally there is the impact of the public authorities which affect the degree of autonomy of the school, the transparency of the information provided by the school, the degree of homogeneity of the school and its location. Together these factors contribute 35.7% to the R-square. It appears therefore that, once genetic factors are excluded, parents and public authorities have quite a similar impact on the efficiency with which a student is able to transform the discretionary inputs at his/her disposal into test scores. These discretionary inputs are respectively the educational means available at home, the inputs of the school (the pedagogical characteristics of school and the physical and human capital available at school) and the time inputs of the student (time devoted to formal as well as informal learning) while, as mentioned previously, the test scores cover four domains, reading, mathematics, science and problem solving.

2) Comparing results for the five South American countries analyzed:

In this section we apply our approach separately to each of the five Latin American countries for which the PISA data were available but, as mentioned previously, for four of them (Brazil, Chile, Mexico and Uruguay) answers to the questionnaire of the parents were not available. We therefore worked with a somehow more limited set of variables and in order to make inter-country comparisons we repeated the analysis for Colombia with this more limited set of variables.

Here also the efficiency analysis[23] involved five inputs. We directly give the results of the OLS regression where the dependent variable is the efficiency value for each individual. The explanatory variables are the 13 determinants of this efficiency, eight of which have been obtained via correspondence analysis while the five others are the gender of the student and the four dummy variables describing the location of the school.

Tables 5-1 to 5-5 in Appendix 5 give the results of this analysis of the determinants of efficiency for each of the five countries. Here also in order to better understand which of these variables has an important impact we implemented a Shapley decomposition.

[23] The results of this analysis may be obtained upon request from the authors.

The results of this investigation are given, separately for each country, in Table 9 to 13.

Here are the main findings of this analysis. For Brazil (see, table 4) it appears that the most important determinants are the location of the school which contribute 36.3% to the R-square, the type of school funding (a contribution of 29.0%), the self-rated ability of the student (a contribution of 16.7%) and the material wealth of the parents (a contribution of 8.0%).

For Chile (see, table 5) the most important contributions are the following ones: the degree of autonomy of the school (30.9%), the human capital of the parents (24.0%), the location of the school (16.1%) and the self-rated ability of the students (13.1%).

For Colombia (table 6) the most important (marginal) contributions are, in order of importance, the gender (31.6%), the location of the school (22.3%), the human capital of the parents (21.3%) and the self-rated ability of the students (7.7%).

For Mexico (table 7) the determinants which have an important marginal contribution to the R-square of the regression are the following ones: the human capital of the parents (33.1%), the location of the school (27.4%), the gender of the student (12.4%), the transparency of the information available at school (7.3%) and the self-rated ability of the students (5.5%).

Finally for Uruguay (table 8) the important determinants are the human capital of the parents (marginal contribution of 39.3%), the self-rated ability of the students (21.8%), the degree of autonomy of the school (12.4%), the type of school funding (10.2%) and the importance of learning in the eyes of the student (7.2%).

The role played by these variables may be summarized as follows. Although it is certainly questionable to try to make international comparisons, we may however note that for all five countries examined, some determinants seem to always appear as being very important. First of all the self-rated ability of the students was classified in all five countries as one of the four or five most important factors. The location of the school is even usually a more important determinant but it does not appear to play an important role in Uruguay, probably because this is a smaller country, eventually with a more homogenous population. Then the human capital of the parents is for three countries among the three most important determinants. Finally the type of school funding and the degree of autonomy of the school were also for two countries one of the most important factors.

We may also try to combine the various determinants into the following three broad categories:

- genetic factors which include the gender of the student and his/her self-rated ability
- the social background of the student which include the parents' human and capital wealth and eventually the importance of learning in the eyes of the student
- the role of the public authorities which determine the location of the school, the type of school funding, the degree of autonomy of the school, the transparency of information at the school and the degree of homogeneity of the school.

Table 14 summarize these findings. Note first that the social background plays the highest role in Uruguay (49.5%) and the smallest in Brazil (14.3%). For the impact of public authorities the highest percentage is observed in Brazil (68.6%) and the smallest in Uruguay (28.5%). Finally for genetic factors the smallest impact is observed in Chile (14.5%) and the highest in Colombia (39.3%). Such a classification is however clearly very questionable because, for example, the relative importance of so-called "genetic factors" is very important for Colombia because gender differences there are extremely high (gender explains 39.3% of the R-square of the regression). Similarly the role of public authorities is particularly important in Brazil and Mexico because the location of the school plays a central role there. Note that the relative importance of public authorities in Chile seems to be due to the high impact in this country of the degree of autonomy of the school.

3) Estimating a common regression for all the countries:

Given the large differences that exist between the countries in the relative contribution of the various determinants of efficiency to the R-square of the regression we have also estimated a regression including the observations for all five Latin American countries. The determinants were the same as previously but this time we also added a dummy variable for the countries. The results of this estimation are given in Table 15. It appears that, ceteris paribus, efficiency is the lowest for Chile and Brazil and the highest for Mexico, Colombia and Uruguay.

Then, on the basis of this regression, we computed again the contribution of the various determinants, included the country dummies, to the R-square of the regression whose results are given in Table 15. These contributions are presented in Table 16 which shows clearly that 41% of the R-square is a consequence of differences between the countries. The second most important variable is the location of the school which contributes 22.5% to the R-square. Two other variables play a relatively important role, the self-rated ability of the student (a contribution of 14.1% to the R-square) and the gender of the students (a contribution of 8.6% to the R-square). Variables related to the characteristics of the school (type of school funding, degree of autonomy of the school, transparency of the information available at school and degree of homogeneity of the school) have together a contribution of 9.9% to the R-square while variables measuring the human or physical capital of the parents contribute only 3.3% to the R-square.

V) Concluding Comments

This paper aimed at taking somehow a new look at the determinants of cognitive ability in five Latina American countries. For each country examined educational production function were estimated by combining efficiency and correspondence analysis. More precisely, using data from the 2006 PISA surveys for Brazil, Chile, Colombia, Mexico and Uruguay, correspondence analysis allowed us to take into account the numerous pieces of information provided by the survey via the questionnaire filled respectively by the parents, the students and the schools.

A first distinction was made between discretionary and non discretionary inputs and then in each of these two cases variables were aggregated in a few determinants obtained on the basis of the first factor of correspondence analysis.

To measure efficiency we adopted the stochastic production frontier approach at the individual (student) level. We thus measured the efficiency with which students were able to transform discretionary inputs into the average scores they received in four types of tests covering respectively the domains of reading, mathematics, science and problem solving. Once such individual efficiency measures were obtained, we estimated via an OLS regression the impact of the non discretionary inputs on this efficiency. Finally, using the so-called Shapley decomposition, we were able to determine the relative impact on efficiency of each of these non discretionary variables.

In a first stage, on the basis of country-specific regressions, we concluded that the location of the school and the self-rated ability of the student had the greatest contribution to the efficiency with which a student is able to transform the discretionary inputs at his/her disposal into test scores. The human capital of the parents, the degree of autonomy of the school and the type of school funding appeared also to play an important role. There were also important differences between the countries in the relative importance of the different determinants.

In a second stage of the analysis, where a unique regression was estimated for all the countries, we concluded that between countries differences explained 41% of the R-square. The other important variables were respectively the location of the school, the self-rated ability of the student, the characteristics of the school and finally the gender of the student.

Given the importance of the location of the school as well as its characteristics which together contribute approximately one third to the R-square, we can quite safely conclude that the quality of schools in rural areas and small cities as well as the relative importance of such factors as the degree of autonomy of the school, the type of school funding, the degree of transparency of the information available on the school and the degree of its homogeneity, have an important impact on the efficiency with which students transform discretionary inputs into test cores.

Table 1: Estimating for Colombia the impact of non discretionary inputs on the efficiency of transforming discretionary inputs into outputs (test scores)

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.70760	0.09141		
Explanatory Variables:				
constant			0.73555	67.38
Gender of the student	0.51429	0.49980	-0.02993	-3.26
Human capital of the parents	-0.00809	0.51195	0.02089	1.95
Material wealth of the parents	0.06154	0.64556	-0.00361	-0.42
Scientific background of the child when he was 10 years old	-0.14420	0.77158	0.00611	1.06
Information on school governance: School funding	-0.15474	0.78078	-0.00086	-0.13
Information on school governance: Autonomy of school	0.24724	0.57966	-0.01302	-1.50
Information on school governance: Transparency of information	0.01176	0.51592	0.01545	1.67
Information on school governance: Homogeneity of school	-0.01328	0.84300	0.00970	1.80
Location of school: village, hamlet or rural area	0.04762	0.21296	-0.00462	-0.20

Location of school: small town (from 3,000 to 15,000 people)	0.51429	0.49980	-0.02993	-3.26
Location of school: town (15,000 to 100,000 people)	-0.00809	0.51195	0.02089	1.95
Location of school: city (100,000 to 1000,000 people)	0.06154	0.64556	-0.00361	-0.42
Importance of learning efforts in the eyes of the student	-0.14420	0.77158	0.00611	1.06
Self-rated ability of student	-0.15474	0.78078	-0.00086	-0.13
ICT (Information and Communication Technology) familiarity of student	0.24724	0.57966	-0.01302	-1.50
Importance of science at home (answers given by the parents)	0.01176	0.51592	0.01545	1.67

Note: R-square: 0.11649; Adjusted R-Square: 0.08141; Number of observations: 420

Table 2: Estimating for Colombia the impact of a limited set of non discretionary inputs on the efficiency of transforming discretionary inputs into test scores (with a smaller number of explanatory variables)

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.70760	0.09141		
Explanatory Variables:				
constant			0.73456	75.15
Gender of the student	0.51429	0.49980	-0.02930	-3.30
Human capital of the parents	-0.00809	0.51195	0.01928	2.01
Scientific background of the child when he was 10 years old	-0.14420	0.77158	0.00605	1.05
Information on school governance: Autonomy of school	0.24724	0.57966	-0.01248	-1.55
Information on school governance: Transparency of information	0.01176	0.51592	0.01573	1.74
Information on school governance: Homogeneity of school	-0.01328	0.84300	0.00956	1.80
Location of school: village, hamlet or rural area	0.04762	0.21296	-0.00350	-0.16

Location of school: small town (from 3,000 to 15,000 people)	0.20714	0.40526	-0.03164	-2.27
Location of school: town (15,000 to 100,000 people)	0.23333	0.42295	-0.01018	-0.83
Location of school: city (100,000 to 1000,000 people)	0.21190	0.40866	0.00672	0.53
Self-rated ability of student	0.00257	0.80119	-0.01025	-1.79
Importance of science at home (answers given by the parents)	0.01069	0.63731	-0.01272	-1.80

Note: R-square: 0.11600; Adjusted R-Square: 0.08994; Number of observations: 420

Table 3: Results of the Shapley Decomposition Procedure for Colombia

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 2	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 2
Gender of the student	0.0190	16.3
Human capital of the parents	0.0223	19.2
Scientific background of the child when he was 10 years old	0.0058	5.0
Information on school governance: Autonomy of school	0.0100	8.6
Information on school governance: Transparency of information	0.0035	3.0
Information on school governance: Homogeneity of school	0.0053	4.6
Location of school	0.0226	19.5
Self-rated ability of student	0.0123	10.6
Importance of science at home (answers given by the parents)	0.0154	13.2
Total	0.1160	100.0

Table 4: Results of the Shapley Decomposition Procedure for Brazil

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 1	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 1
Gender of the student	0.0003	0.5
Human capital of the parents	0.0020	3.5
Material wealth of parents	0.0046	8.0
Information on school governance: School funding	0.0166	29.0
Information on school governance: Autonomy of school	0.0015	2.6
Information on school governance: Transparency of information	0.0002	0.3
Information on school governance: Homogeneity of school	0.0002	0.4
Location of school	0.0208	36.3
Importance of learning in eyes of student	0.0016	2.8
Self-rated ability of student	0.0096	16.7
Total	0.0575	100.0

Table 5: Results of the Shapley Decomposition Procedure for Chile

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 2	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 2
Gender of the student	0.0058	3.5
Human capital of the parents	0.0393	23.6
Material wealth of parents	0.0049	2.9
Information on school governance: School funding	0.0073	4.4
Information on school governance: Autonomy of school	0.0486	29.3
Information on school governance: Transparency of information	0.0015	0.9
Information on school governance: Homogeneity of school	0.0092	5.5
Location of school	0.0252	15.2
Importance of learning in eyes of student	0.0062	3.8
Self-rated ability of student	0.0183	11.0
Total	0.1662	100.0

Table 6: Results of the Shapley Decomposition Procedure for Colombia

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 3	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 3
Gender of the student	0.0405	31.6
Human capital of the parents	0.0272	21.3
Material wealth of parents	0.0021	1.6
Information on school governance: School funding	0.0070	5.5
Information on school governance: Autonomy of school	0.0031	2.4
Information on school governance: Transparency of information	0.0028	2.2
Information on school governance: Homogeneity of school	0.0014	1.1
Location of school	0.0285	22.3
Importance of learning in eyes of student	0.0054	4.2
Self-rated ability of student	0.0098	7.7
Total	0.1280	100.0

Table 7: Results of the Shapley Decomposition Procedure for Mexico

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 4	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 4
Gender of the student	0.0100	12.4
Human capital of the parents	0.0268	33.1
Material wealth of parents	0.0022	2.8
Information on school governance: School funding	0.0016	2.0
Information on school governance: Autonomy of school	0.0013	1.7
Information on school governance: Transparency of information	0.0059	7.3
Information on school governance: Homogeneity of school	0.0045	5.5
Location of school	0.0221	27.4
Importance of learning in eyes of student	0.0020	2.4
Self-rated ability of student	0.0044	5.5
Total	0.0808	100.0

Table 8: Results of the Shapley Decomposition Procedure for Uruguay

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 5	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 5
Gender of the student	0.0005	0.3
Human capital of the parents	0.0557	39.3
Material wealth of parents	0.0042	3.0
Information on school governance: School funding	0.0144	10.2
Information on school governance: Autonomy of school	0.0176	12.4
Information on school governance: Transparency of information	0.0007	0.5
Information on school governance: Homogeneity of school	0.0028	2.0
Location of school	0.0048	3.4
Importance of learning in eyes of student	0.0102	7.2
Self-rated ability of student	0.0309	21.8
Total	0.1418	100.0

Table 9: Attempting to estimate the relative importance (in percentage terms) of the social background, genetic factors and the role of public authorities

Country	Impact of Social Background	Impact of Public Autorities	Impact of Genetic Factors
Brazil	14.3	68.6	17.2
Chile	30.3	55.3	14.5
Colombia	27.1	33.5	39.3
Mexico	38.3	43.9	17.9
Uruguay	49.5	28.5	22.1

Table 10: Estimating together, for all five Latin American countries, the impact of a limited set of non discretionary inputs on the efficiency of transforming discretionary inputs into test scores

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.67090	0.09096		
Explanatory Variables:				
Constant			0.66026	161.25
Gender of the student	0.53306	0.49891	-0.01600	-7.70
Human capital of the parents	0.00592	0.54809	-0.00741	-3.77
Material welath of parents	-0.04119	0.84666	0.00122	0.92
Information on school governance: **School funding**	-0.02888	1.01497	-0.00556	-5.08
Information on school governance: **Autonomy of school**	0.06325	0.53948	-0.00362	-1.81
Information on school governance: **Transparency of information**	-0.02019	0.60847	0.00725	4.24
Information on school governance: **Homogeneity of school**	0.01285	0.94880	0.00333	3.04
Location of school: village, hamlet or rural area	0.05182	0.22166	-0.03764	-7.25

Location of school: small town (from 3,000 to 15,000 people)	0.17394	0.37906	-0.01775	-5.14
Location of school: town (15,000 to 100,000 people)	0.24743	0.43152	-0.00394	-1.25
Location of school: city (100,000 to 1000,000 people)	0.30967	0.46236	0.00989	3.18
Importance of learning in eyes of student	-0.00023	0.73048	-0.00207	-1.46
Self-rated ability of student	0.00691	0.79358	-0.01167	-8.90
Dummy variable for Chile	0.09780	0.29705	-0.02005	-4.37
Dummy variable for Colombia	0.12378	0.32934	0.02826	6.50
Dummy variable for Mexico	0.52667	0.49929	0.03421	9.80
Dummy variable for Uruguay	0.14226	0.34932	0.02044	4.78

Note: R-square: 0.08088 ; Adjusted R-Square: 0.07870;
Number of observations: 7198

Table 11: Results of the Shapley Decomposition Procedure on the Basis of the Results of Table 10

Explanatory Variables	Marginal Contribution of the variable to the R-Square of the regression in Table 12	Marginal Contribution (in percentage terms) of the variable to the R-Square of the regression in Table 12
Gender of the student	0.0070	8.6
Human capital of the parents	0.0025	3.0
Material wealth of parents	0.0002	0.3
Information on school governance: School funding	0.0027	3.3
Information on school governance: Autonomy of school	0.0006	0.7
Information on school governance: Transparency of information	0.0028	3.5
Information on school governance: Homogeneity of school	0.0019	2.4
Location of school	0.0182	22.5
Importance of learning in eyes of student	0.0004	0.5
Self-rated ability of student	0.0114	14.1
Country Dummy Variables	0.0332	41.0
Total	0.0809	100.0

References

Afonso, A,. and M. St. Aubyn, "Cross-Country Efficiency of Secondary Education Provision. A Semi-parametric Analysis with Nondiscretionary Inputs," European Central Bank, Working Paper Series, No. 494, 2005.

Aigner, D., C. A. K. Lovell and P. Schmidt, “Formulation and Estimation of Stochastic Frontier Production Function Models,” *Journal of Econometrics* 6 (1977): 21- 37.

Aigner, D., C. A. K. Lovell and P. Schmidt, “Formulation and Estimation of Stochastic Frontier Production Function Models,” *Journal of Econometrics* 6 (1977): 21- 37.

Anderson, G., I. Crawford and A. Leicester, "Efficiency Analysis and the Lower Convex Hull Approach," Chapter 10 in N. Kakwani and J. Silber editors, *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave-Macmillan, 2008.

Asselin, L.-M. and Vu Tuan Anh "Multidimensional Poverty and Multiple Correspondence Analysis," Chapter 5 in N. Kakwani and J. Silber editors, *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave-Macmillan, 2008.

Athanassopoulos, A. D. and E. Shale, "Assessing the comparative efficiency of higher education institutions in the UK by means of data envelopment analysis," *Education Economics*, 5(2) (1997): 117-134.

Barro, R., "Human Capital and Growth," *American Economic Review* 91(2) (2001): 12-17.

Bates, J. M., "Measuring predetermined socioecoomic 'inputs' when assessing the efficiency of educational outputs," *Applied Economics*, 29(1997): 85-93.

Battese, G. E. and G. S. Corra, “Estimation of a Production Frontier Model: With Application to the Pastoral Zones of Eastern Australia,” *Australian Journal of Agricultural Economics* 21 (1977): 169-179.

Behrman, J. R. and M. R. Rosenzweig, “Does Increasing Women’s Schooling Raise the Schooling of the Next Generation?” *American Economic Review* 92:1(March 2002): 323-334.

Behrman, J. R. and M. R. Rosenzweig, “Does Increasing Women’s Schooling Raise the Schooling of the Next Generation? – Reply,” *American Economic Review* 95:5 (December 2005): 1745-1751.

Behrman, J. R., J. Hoddinott, J. A. Maluccio, E. Soler-Hampejsek, E. L. Behrman, R. Martorell, M. Ramirez and A. D. Stein, 2008, “What Determines Adult Cognitive Skills? Impacts of Pre-School, School-Years and Post-School Experiences in Guatemala”, Washington, DC: IFPRI Discussion Paper No. 826.

Benzécri, J.-P. and F. Benzécri, *Pratique de L'Analyse des Données, I, Analyse des Correspondances, Exposé Elémentaire,* Paris: Dunod Bordas, 1980.

Berrebi, Z. M. and J. Silber, "Health and Development: Socio-Economic Determinants of Mortality Structure," *Social Science and Medicine*, 15C (1981): 31-39.

Bessent, A. M., W. Bessent, J. Kennington and B. Reagan, "An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District," *Management Science* 28(1982): 1335-1366.

Black, S. E., P. J. Devereux and K. G. Salvanes, “Why the Apple Doesn’t Fall Far: Understanding Intergenerational Transmission of Human Capital,” *American Economic Review* 95:1 (March 2005): 437-449.

Chantreuil, F. and A. Trannoy, "Inequality Decomposition Values: The Trade-Off Between Marginality and Consistency," THEMA Discussion Paper, Université de Cergy-Pontoise, 1999.

Coelli, T. J., "A Computer program for Frontier Production Function Estimation: Version Frontier 2.0," *Economics Letters* 39: 29-32.

Coelli, T., D.S. Prasada Rao and G. E. Battese, *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, Boston, 1998.

D'Ambrosio, C., J. Deutsch and J. Silber, ""Multidimensional Approaches to Poverty Measurement: An Empirical Analysis of Poverty in Belgium, France, Germany, Italy and Spain, based on the European Panel," *Applied Economics*, forthcoming.

Deaton, A., "The Distance Function in Consumer Behavior with Application to Index Numbers and Optimal Taxation," *Review of Economic Studies* 46 (1979): 391-405.

Deaton, A. and J. Muellbauer, *Economics and Consumer Behavior*, Cambridge University Press, Cambridge, 1980.

Deller, S. C. and E. R. Rudnicki, "Production efficiency in elementary education: The case of Maine public schools, " *Economics of Education Review*, 12 (1993): 45-57.

Deutsch, J., X. Ramos and J. Silber, "Poverty and Inequality of Standard of Living and Quality of Life in Great Britain," chapter 7 in *Advances in Quality-of-Life Theory and Research*, J. Sirgy, D. Rahtz and A.C. Samli, editors, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003, pp. 99-128.

Deutsch, J. and J. Silber, “Religion, Standard of Living and the Quality of Life,” *Contemporary Jewry*, 20 (1999):119-137.

Deutsch, J. and J. Silber, “Ethnic Origin and Multidimensional Relative Poverty in Israel: A Study Based on the 1995 Israeli Census,” in *The Economics of Immigration and Social Diversity, Research in Labor Economics*, S. W. Polachek, C. Chiswick and H. Rapoport, editors, 24 (2006): 235-264.

Dolton, P., O. D. Marcenaro and L. Navarro, "The Effective Use of Student Time: A Stochastic Frontier Production Function Case Study," mimeo, 2001.

Duncombe, W., J. Miner and J. Ruggiero, "Empirical evaluation of bureaucratic models of inefficiency," *Public Choice*, 93(1) (1997): 1-18.

Farrell, M. J., “The Measurement of productive Efficiency,” *Journal of the Royal Statistical Society,* Series A, CXX(1957): 253-290.

Glewwe, P. and M. Kremer, "Schools, Teachers and Education Outcomes i*n Developing Countries," Chapter 16 in* Handbook of the Economics of *Education*, edited by E. Hanushek and F. Welch, Vol. 2, 2006: 945-1017.

Glewwe, P., N. Ilias and M. Kremer, 2003, "Teacher Incentives," N.B.E.R. working paper 9671.

Glewwe, P., M. Kremer and S. Moulin, 2007, "Many Children Left Behind? Textbooks and Test Scores in Kenya," N.B.E.R. working paper W13300.

Green, W. H., LIMDEP Version 6.0: User’s Manual and Reference Guide, Econometric Software Inc., New York, 1992.

Grosskopf, S., K. Hayes, L. Taylor and W. Weber, "On the determinants of school district efficiency: Competition and monitoring," *Journal of Urban Economics*, 49 (2001): 453-478.

Gstach, D., A. Somers and S. Warning, "Output specific efficiencies: the case of UK private secondar schools," Working Paper No. 84, Vienna University of Economics, Vienna, 2003.

Hanushek, E. "The economics of schooling: Production and efficiency in public schools," *Journal of Economic Literature* XXIV (1986) (3): 1141–1177.

Hanushek, E. and D. Kimko, "Schooling, labor force quality and economic growth," *American Economic Review*, 90(5) (2000): 1184-1208.

Heckman, J. J., J. Stixrud and S. Urzua, "The Effect of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," N. B. E. R. Working Paper 12006, 2006.

Israeli, O., "A Shapley-based decomposition of the R-Square of a linear regression," **.** *Journal **of** Economic Inequality***,** 5(2) (2007); 199-212.

Johnes, J., "Data envelopment analysis and its application to the measurement of efficiency in higher education," *Economics of Education Review*, 25 (2006): 273-288.

Johnes, G. and J. Johnes, "Measuring the research performance of UK economics department: An application of data envelopment analysis," *Oxford Economic Papers* 45(1993): 332-347.

Johnes, J. and G. Johnes, "Research funding and performance in UK university departments of economics: a frontier analysis," *Economics of Education Review*, 14(1995): 301-314.

Johnson, R. A. and D. W. Wichern, 1999, *Applied Multivariate Statistical Analysis*, Prentice Hall.

Kakwani, N. and J. Silber, editors, *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave-Macmillan, 2008.

Lovell, C. A. K., S. Richardson, P. Travers and L. Wood, "Resources and Functionings: A New View of Inequality in Australia," in *Models and Measurement of Welfare and Inequality*, W. Eichhorn, editor, Springer-Verlag, Heidelberg, 1994.

Madden, G., S. Savage and S. Kemp, "Measuring public sector efficiency: A study of economics departments at Australian universities," *Education Economics*, 5 (1997): 153-168.

Mancebon, M. J. and E. Bandrès, "Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results," *Education Economics*, 7(2) (1999): 131-152.

Meeusen, W. and J. van den Broeck, "Efficiency Estimates from Cobb-Douglas Production Functions with Composed Error," *International Economic Review* 18(1977): 435-444.

OECD Programme for International Student Assessment (PISA), available at http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html

Plug E. and W. P. M. Vijverberg, "Schooling, Family Background and Adoption: Is it Nature or Nurture?," *Journal of Political Economy* 111 (2003): 611-41.

Plug, E. , "Estimating the Effect of Mother's Schooling on Children's Schooling Using a Sample of Adoptees," *American Economic Review* 94:1 (March 2004): 358-368.

Portela, M. C. and A. S. Camanho, "Performance Assessment of Portuguese Secondary Schools," mimeo, 2007.

Pritchett, L., "Where Has All the Education Gone," *World Bank Economic Review*, 15(3) (2001): 367-391.

Pritchett, L. and D. Filmer, "What production functions really show: A positive theory of education expenditures," *Economics of Education Review*, 18 (1999): 223-239.

Ramos, X. and J. Silber, "On the Application of Efficiency Analysis to the Study of the Dimensions of Human Development," *Review of Income and Wealth*, 51(2) (2005): 285-309.

Richmond, J. "Estimating the efficiency of production," *International Economic Review*, **15**(2) (1974): 515-21.

Ruggiero, J., "Efficiency of Educational Production: An Analysis of New York School Districts," Review of Economics and Statisticsa, 78(3), (1996): 449-509.

Sastre, M. and A. Trannoy, "Shapley Inequality Decomposition by Factor Components: Some Methodological Issues," *Journal of Economics*, Supplement 9 (2002): 51-89.

Shapley, L. S. "A value for n-persons games,". In *Contributions to the Theory of Games II*. A. W. Tucker and R. D. Luce eds., Princeton University Press, 307-317, 1953.

Shorrocks, A. F., "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," mimeo, University of Essex, 1999.

Simar, L. and P. W. Wilson, "Estimation and inference in two-stage semi-parametric models of production process," *Journal of Econometrics* 136(2007): 31-64.

Todd, P. E. and K. I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal*, 113 (2003): F3-F33.

Todd, P. E. and K. I. Wolpin, 2007, “The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps,” *Journal of Human Capital* 1 (Winter 2007): 91-136.

Urzua, S., "Racial Labor Markets Gaps: The Role of Abilities and Schooling Choices," mimeo, 2007.

Waldo, S., "Competition and Public School Efficiency in Sweden – An Empirical Evaluation of Second-Stage Results for Different Models of Non-Discretionary Inputs in Data Envelopment Analysis," mimeo, 2006.

Waldo, S., "On the use of student data in efficiency analysis – Technical efficiency in Swedish upper secondary school," *Economics of Education Review*, 26(2007): 173-185.

Wilson, P. W., "Efficiency in Education Production among PISA Countries, with Emphasis on Transitioning Economies," mimeo, 2005.

Worthington, A. C., "An empirical survey of frontier efficiency measurement techniques in education," *Education Economics*, 9(3) (2001): 245-268.

Appendix 1: Detailed and classified list of the PISA questions that were used[24]

I) Inputs in the education production process:

A1) Educational means available at home:

These inputs were taken from both the parent and the student questionnaires. Here is the list of these inputs that were selected, at the light of the available data..

- Specific educational means available at home:
  - a desk to study at (variable S4)
  - a room of your own (variable S5)
  - a quiet place to study (variable S6)
  - a computer you can use for school work (variable S7)
  - educational software (variable S8)
  - a link to the Internet (variable S9)
  - your own calculator (variable S10)
  - classic literature (variable S11)
  - books of poetry (variable S12)
  - works of art (variable S13)
  - books to help with school work (variable S14)
  - a dictionary (variable S15)
  - a DVD or VCR player (variable S16)

B) The inputs of the school:

B1) Pedagogical characteristics of school:

- Are students grouped by ability within their classes (for all, some or no subject): variable B6 with 2 categories
- Is your school involved in
  - science clubs (variable B7)
  - science fairs (variable B8)

[24] Questions with an asterisk (*) were available only for Colombia.

- science competitions (variable B9)

- extracurricular science projects (variable B10)

- excursion and field trips (variable B11)

- When learning school science topics at school (this question is taken from the student's questionnaire), how often do the following activities occur?

- students are given opportunities to explain their ideas (variable S18 with 3 categories)

- students spend time in the laboratory doing practical experiments (variable S19 with 3 categories)

- students are required to design how a school science question could be investigated in the laboratory (variable S20 with 3 categories)

- students are asked to apply a school science concept to everyday problems (variable S21 with 3 categories)

- the lessons involve students' opinions about the topic (variable S22 with 3 categories)

- students are asked to draw conclusions from an experiment they have conducted (variable S23 with 3 categories)

- the teacher explains how a school science idea can be applied to a number of different phenomena (variable S24 with 3 categories)

- students are allowed to design their own experiment (variable S25 with 3 categories)

- there is a class debate or discussion (variable S26 with 3 categories)

- experiments are done by the teacher as demonstrations (variable S27 with 3 categories)

- students are given a chance to choose their own investigation (variable S28 with 3 categories)

- the teacher uses school science to help the student understand the world outside school (variable S29 with 3 categories)

- students have discussions about the topics (variable S30 with 3 categories)

- students do experiments by following the instructions of the teacher (variable S31 with 3 categories)

- the teacher clearly explains the relevance of broad science concepts to our lives (variable S32 with 3 categories)

- students are asked to do an investigation to test out their own ideas (variable S33 with 3 categories)

- the teacher uses examples of technological application to show how school science is relevant to society (variable S34 with 3 categories)

B2) Physical and Human Capital available at school:

- Total school enrollment (number of students) of school (variables B12 for boys and B13 for girls)
- How many teachers as a whole are working full time? (variable B15)
- About how many computers are available in the school for instruction? (variable B22)
- About how many computers are connected to the Internet? (variable B23)
- Is your school's capacity to provide instruction hindered by
    - a lack of qualified science teachers (variable B24, 3 categories)
    - a lack of qualified mathematics teachers (variable B25, 3 categories)
    - a lack of qualified language teachers (variable B26, 3 categories)
    - a lack of qualified teachers in other subjects (variable B27, 3 categories)
    - a lack of laboratory technicians (variable B28, 3 categories)
    - a lack of other support personnel (variable B29, 3 categories)
    - a shortage or inadequacy of science laboratory equipment ((variable B30, 3 categories)
    - a shortage or inadequacy of instructional materials (e.g. textbooks): (variable B31, 3 categories)
    - a shortage or inadequacy of computers for instruction (variable B32, 3 categories)
    - a lack or inadequacy of Internet connectivity (variable B33, 3 categories)
    - a shortage or inadequacy of computer software for instruction (variable B34, 3 categories)
    - a shortage or inadequacy of library materials (variable B35, 3 categories)
    - a shortage or inadequacy of audio-visual resources (variable B36, 3 categories)

C) The inputs of the student:

C1) Time devoted by the student to informal learning:

- How often do you (possible answers: very often, regularly, sometimes, never or hardly ever)
    - watch TV programs about science (variable S35, 3 categories)
    - borrow or buy books on "broad science" (variable S36, 3 categories)

- visit websites about "broad science" topics (variable S37, 3 categories)
- listen to radio programs about advances in "broad science" (variable S38, 3 categories)
- read "broad science" magazines or science articles in newspapers (variable S39, 3 categories)
- attend a "science club" (variable S40, 3 categories)

C2) Time devoted to formal learning:

- How much time do you typically spend per week studying the following subjects (possible answers: no time, less than 2 hours, 2 to 4 hours, 4 to 6, 6 or more):
- regular lessons in school science at school (variable S41, 4 categories)
- out-of-school-time lessons in school science (variable S42, 4 categories)
- study or homework in school science by yourself (variable S43, 4 categories)
  - mathematics:
    - regular lessons in mathematics at school (variable S44, 4 categories)
    - out-of-school-time lessons in mathematics (variable S45, 4 categories)
    - study or homework in mathematics by yourself (variable S46, 4 categories)
  - language (of the test):
    - regular lessons in this language at school (variable S47, 4 categories)
    - out-of-school-time lessons in this language (variable S48, 4 categories)
    - study or homework in this language by yourself (variable S49, 4 categories)
  - other subjects:
    - regular lessons in other subjects at school (variable S50, 4 categories)
    - out-of-school-time lessons in other subjects (variable S51, 4 categories)
    - study or homework in other subjects by yourself (variable S52, 4 categories)

- What type of out-of-school-time lessons do you attend currently?
  - one to one lessons with a teacher who teaches also at your school (variable S53)
  - one to one lesson with a teacher who does not teach at your school (variable S54)
  - lessons in small groups (less than 8 students) with a teacher who teaches at your school (variable S55)
  - lessons in small groups (less than 8 students) with a teacher who does not teach at your school (variable S56)
  - lessons in large groups (8 students or more) with a teacher who teaches at your school (variable S57)

- lessons in large groups (8 students or more) with a teacher who does not teach at your school (variable S58)

II) Factors affecting the efficiency of inputs use in producing the output:

1) Gender of the student (variable z1)

2) Human capital of the parents:

- Age of the father (younger than 36 year, 36 to 40 years, 41 to 45 years, 46 to 50 years, 51 years and older) when the child was born (variables z21 to z24)
- Age of the mother (younger than 36 year, 36 to 40 years, 41 to 45 years, 46 to 50 years, 51 years and older) when the child was born (variables z25 to z28)
- *Main job of the father (one digit classification: variables z29 to z217)
- *Main job of the mother (one digit classification: variables z218 to z226)
- *Highest schooling level completed by the father (variables z227 to z229)
- *Highest schooling level completed by the mother (variables z230 to z232)
- *Country of birth of father (in the country or abroad: variable z233)
- *Country of birth of mother (in the country or abroad: variable z234)
- *Language spoken at home most of the time (language of the test or not, variable z235)

3) Material wealth of parents:

- What is the annual household income (on a scale from 1 to 6) (variables z31 to z35)
- How many of the following items are at home?
    - a dishwasher (variable z36: yes or no)
    - cellular phones (variables z37 to z39: none, 1 ,2, 3 or more)
    - televisions (variables z310 to z312: none, 1 ,2, 3 or more)
    - *computers (variables z313 to z315: none, 1 ,2, 3 or more)
    - *cars (variables z316 to z318: none, 1 ,2, 3 or more)

4) Scientific background of the child when he was 10 years old:

- *Thinking back to when your child was about 10 years old, how often would your child have watched TV programmes about science? (the possible answers to this question are: very often, regularly, sometimes, never) (variables z41 to z43)

- *Thinking back to when your child was about 10 years old, how often would your child have read books on scientific discoveries? (the possible answers to this question are: very often, regularly, sometimes, never) (variables z44 to z46)
- *Thinking back to when your child was about 10 years old, how often would your child have watched, read or listened to science fiction? (the possible answers to this question are: very often, regularly, sometimes, never) (variables z47 to z49)
- *Thinking back to when your child was about 10 years old, how often would your child have visited websites about science topics? (the possible answers to this question are: very often, regularly, sometimes, never) (variables z410 to z412)
- *Thinking back to when your child was about 10 years old, how often would your child have attended a science club? (the possible answers to this question are: very often, regularly, sometimes, never) (variables z413 to z415)

5) Information on school governance:

5-1: School funding

- Is the school a private or a public school? (variable z51)
- What percentage of the school's total funding for a typical school year comes from the government (local, regional, state or national) (variable z52 is equal to 1 if the percentage is higher than 50%, to 0 otherwise)
- What percentage of the school's total funding for a typical school year comes from student fees or school charges paid by parents (variable z53 is equal to 1 if the percentage is higher than 50%, to 0 otherwise)
- What percentage of the school's total funding for a typical school year comes from benefactors, donations, bequests, sponsorships, parental fund raising? (variable z54 is equal to 1 if the percentage is higher than 20%, to 0 otherwise)
- What percentage of the school's total funding for a typical school year comes from other sources (variable z55 is equal to 1 if the percentage is higher than 20%, to 0 otherwise)

5-2: Autonomy of school:

Regarding your school which of the following categories (principal or teachers, school governing board, regional or local education authority, national education authority) has a considerable responsibility for

- selecting teachers for hire: (binary variable: principal or other, variable z61)

- firing teachers: (binary variable: principal or other, variable z62)

- establishing teachers' starting salaries: (binary variable: principal or other, variable z63)

- determining teachers' salary increases: (binary variable: principal or other, variable z64)

- formulating the school's budget: (binary variable: principal or other, variable z65)

- deciding on budget allocations within the school: (binary variable: principal or other, variable z66)

- establishing student disciplinary policies: (binary variable: principal or other, variable z67)

- establishing student assessment policies: (binary variable: principal or other, variable z68)

- approving students for admission to the school: (binary variable: principal or other, variable z69)

- choosing which textbooks are used: (binary variable: principal or other, variable z610)

- determining course content: (binary variable: principal or other, variable z611)

- *deciding which courses are offered: (binary variable: principal or other, variable z612)

5-3: Transparency of information

- Do you feel (constant, minor, no) pressure from parents concerning the need of high academic standards? (variables z71 to z72)
- Are achievement data in your school (yes or no)
    - posted publicly (variable z73)
    - used in evaluating the principal's performance (variable z74)
    - used in evaluating the teachers' performance (variable z75)
    - used in decisions about instructional resource allocation to the school (variable z76)
    - tracked over time by an administrative authority (variable z77)
- Are there (two or more, one, no) schools in your area that compete for the students? (variables z78 to z79)
- Does your school provide information to parents of students on their child academic performance relative to other students of similar grade in your school? (yes or no: variable z710)

- Does your school provide information to parents of students on their child academic performance relative to other students of similar grade at the national or regional level? (yes or no: variable z711)

- Does your school provide information to parents of students on the academic performance of students of your child's age in your school compared to other schools? (yes or no: variable z712)

5-4: Homogeneity of school:

- How much consideration is given to the following factors when students are admitted to school (the possible answers are: prerequisite, high priority, considered, not considered)
  - residence in a particular area (variables z81 to z83)
  - student's academic record (variables z84 to z86)
  - parents' endorsement of instructional or religious philosophy of school (variables z87 to z89)
  - student's need or desire for a special program (variables z810 to z812)
  - attendance of other family members at the school (in the past or present) (variables z813 to z815)

5) Location of school:

- Which of the following best describes the community in which your school is located: village/hamlet/rural area, small town (from 3,000 to 15,000 people), town (15,000 to 100,000 people), city (100,000 to 1000,000 people), large city (more than a million) (variables z91 to z94)

6) Importance of learning efforts in eyes of student:

- How important do you think it is for you to do well (possible answers: very important, important, of little importance, not important at all) in:
  - school science subject (variables z101 to z103)
  - mathematics (variables z104 to z105; note that one of the possible answers was not chosen by anybody so that we include only two and not three categories)
  - test language (variables z106 to z108)

7) Self-rated ability of student:

How much do you agree with the statements below (possible answers: strongly agree, agree, disagree, strongly disagree):

- learning advanced school science topics would be easy for me (variables z111 to z113)
- I can usually give good answers to test questions on school science topics (variables z114 to z116)
- I learn school science topics quickly (variables z117 to z119)
- When I am taught school science, I can understand the concepts very well (variables z1110 to z1112)
- I can easily understand new ideas in school science (variables z1113 to z1115)

8) ICT familiarity of student:

- *Have you ever used a computer (yes or no: variable z121)
- *How long have you been using a computer? (less than one year, one to three years, three to five years, five years or more) (variables z122 to z124)
- *How often do you use a computer at the following places (possible answers: almost every day, once or twice a week, a few times a month, once a month ro less, never):
    - at home (variables z125 to z128)
    - at school (variables z129 to z1212)
    - at other places (variables z1213 to z1216)
- *How often do you use computers for the following reasons (possible answers: almost every day, once or twice a week, a few times a month, once a month or less, never):
    - browse the internet for information about people, things, ideas (variables z1217 to z1220)
    - play games (variables z1221 to z1224)
    - write documents (e.g. with Word or Word Perfect) (variables z1225 to z1228)
    - use the internet to collaborate with a group or team (variables z1229 to z1232)
    - use spreadsheets (e.g. Lotus or Excel) (variables z1233 to z1236)
    - download software from the internet (including games) (variables z1237 to z1240)
    - drawing, painting or using graphic programs (variables z1241 to z1244)

- use educational software such as Mathematics programs (variables z1245 to z1248)

- download music from the Internet (variables z1249 to z1252)

- writing computer programs (variables z1253 to z1256)

- for communication (e.g. E-mail or "chat rooms") (variables z1257 to z1260)

- *How well can you do each of these tasks on a computer (possible answers: I can do this very well by myself, I can do this with the help from someone, I know what this means but I cannot do it, I do not know what this means):

- chat online ((variables z1261 to z1263)

- use software to find and get rid of computer viruses (variables z1264 to z1266)

- edit digital photographs or other graphic images (variables z1267 to z1269)

- create a database (e.g. use Microsoft Access) (variables z1270 to z1272)

- copy data to a CD (e.g. make a music CD) (variables z1273 to z1275)

- move files from one place to another on a computer (variables z1276 to z1278)

- search the internet for information (variables z1279 to z1281)

- download files or programs from the Internet (variables z1282 to z1284)

- attach a file to an E-mail message (variables z1285 to z1287)

- use a word processor (e.g. to write an essay for school) (variables z1288 to z1290)

- use a spreadsheet to plot a graph (variables z1291 to z1293)

- create a presentation (e.g. using Microsoft Poer Point) (variables z1294 to z1296)

- download music form the internet (variables z1297 to z1299)

- create a multimedia presentation (with sound, pictures, video) (variables z12100 to z12102)

- write and send E-mails (variables z12103to z12105)

- construct a web page (variables z12106 to z12108)

9) Importance of science at home (answers given by the parents):

- How much do you agree with the following statement: it is important to have good scientific knowledge and skills in order to get any good job in today's world (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z131 to z133)
- How much do you agree with the following statement: employers generally appreciate strong scientific knowledge and skills among their employees (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z134 to z136)
- How much do you agree with the following statement: most jobs today require some scientific knowledge and skills (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z137 to z139)
- How much do you agree with the following statement: it is an advantage in the job market to have good scientific knowledge and skills (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z1310 since there were two possible answers that nobody selected)
- How much do you agree with the following statement: advances in "broad science and technology" usually improve people's living condition (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z1311 to z1312; note that here again there was nobody selecting one of the categories)
- How much do you agree with the following statement: "broad science" is important for helping us understand the natural world (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z1313 to z1314; as here again one of the categories waqs not selected by anybody)
- How much do you agree with the following statement: "broad science" usually is valuable to society (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z1315 to z1316, with here again one category not selected by anybody)
- How much do you agree with the following statement: advances in "broad science and technology" usually bring social benefits (the possible answers to this question are: strongly agree, agree, disagree, strongly disagree) (variables z1317 to z1318, with here again one category not selectedby anybody)

Appendix 2: On Correspondence Analysis

Correspondence analysis (CA) was originally introduced by Benzecri and Benzecri (1980). It is strongly related to principal components analysis (PCA) but while PCA assumes that the variables are quantitative, CA has been designed to deal with categorical variables. More precisely CA offers a multidimensional representation of the association between the row and column categories of a two-way contingency table. In short the goal of CA is to find scores for both the row and column categories on a small number of dimensions (axes) that will account for the greatest proportion of the *chi²* measuring the association between the row and column categories. There is thus a clear parallelism between CA and PCA, the main difference being that PCA[25] accounts for the maximum *variance*. A clear presentation of CA is given in Asselin and Vu Tuan Anh (2008), chapter 5 in Kakwani and Silber (2008).

Let us first recall what the main features of PCA. It is in fact a data reduction technique that consists of building a sequence of orthogonal and normalized linear combinations of the K primary indicators that will exhaust the variability of the set primary indicators. These orthogonal linear combinations are evidently latent variables and usually called "components". In PCA the first component has the greatest variance and all subsequent components have decreasing variances.

Let *N* be the size of the population, *K* the number of indicators I_k . The first component F^1 may be expressed for observation *i* as

$$F_i^1 = \sum_{k=1}^{K} \omega_k^1 I_i^{*k} \ .$$

where I^{*k} refers to the standardized primary indicator I^k. Note that ω_k^1 is the (first) factor score coefficient for indicator *k*. It turns out that the scores ω_k^1 are in fact the multiple regression coefficients between the component F^1 and the standardized primary indicators I^{*k} . It is very important to understand that PCA has some limitations, of which the most important is probably the fact that PCA has been developed for *quantitative* variables.

It is therefore better not to use PCA when some of the variables are of a qualitative nature. (Multiple) Correspondence Analysis (MCA) is in fact the data reduction technique that should be used in the presence of categorical variables

[25] For an illustration of the use of PCA, see, for example, Berrebi and Silber,1981.

Let us therefore assume now that the *K* primary indicators are categorical ordinal and that the indicator I^k has J^k categories. Note that if some of the variables of interest are quantitative, it is always possible to transform them into a finite number of categories. To each primary indicator I^k we therefore associate the set of J^k binary variables that can only take the value 0 or 1.

Let us now call $X(N, J)$ the matrix corresponding to the *N* observations on the *K* indicators which are now decomposed into J^k variables. Note that $J = \sum_{k=1}^{K} J^k$ represents now the total number of categories. Call N_j the absolute frequency of category ***j***. Clearly N_j is equal to the sum of column *j* of the matrix *X*. Let $N_{..}$ refer to the sum of all the $(N \ by \ K)$ elements of the matrix *X*. Let also f_j be the relative frequency $(N_j / N_{..})$, f^i be the sum of the i^{th} line of matrix *X*, f_{ij} be the value of cell (*i*,*j*) and f_j^i be equal to the ratio (f_{ij} / f^i). Finally call $\{f_j^i\}$ the set of all f_j^i's for a given observation *i* (*j* = 1 to *J*). This set will be called the profile of observation *i*.

As stressed previously CA is a PCA process applied to the matrix *X*, but with the χ^2 - metric on row/column profiles, instead of the usual Euclidean metric. This χ^2 - metric is in fact a special case of the Mahalanobis distance developed in the 1930s. This metric defines the distance $d^2(f_j^i, f_j^{i'})$ between two profiles *i* and *i'* as

$$d^2(f_j^i, f_j^{i'}) = \sum_{j=1}^{J} (1/f_j)(f_j^i - f_j^{i'})^2$$

Note that the only difference with the Euclidean metric lies in the term $(1/f_j)$. This term indicates that categories which have a low frequency will receive a higher weight in the computation of distance. As a consequence CA will be overweighting the smaller categories within each primary indicator. It can be shown that

$$\omega_j^{1,k} = \frac{1}{(N_j^k / N)} Cov(F^{1*}, I_j^k)$$

where $\omega_j^{1,k}$ is the score of category j_k on the first (non-normalized) factorial axis, I_j^k is a binary variable taking the value 1 when the population unit belongs to the category j_k, F^{1*} is the normalized score on the first axis and N_j^k is the frequency of the category j_k of indicator *k*.

Ir is also very interesting to note that CA offers a unique duality property since it can be shown that

$$F_1^i = \frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \frac{w_j^{1,k}}{\lambda_1} I_{i,j}^k}{K}$$

where *K* is the number of categorical indicators, J_k is the number of categories for indicator *k*, $w_j^{1,k}$ is the score of category j_k on the first (non normalized) factorial axis, $I_{i,j}^k$ is a binary variable taking the value 1 when unit *i* belongs to category j_k and F_1^i is the (non normalized) score of observation *i* on the first factorial axis[26].

Reciprocally it can be shown that

$$\omega_j^{1,k} = \frac{\sum_{i=1}^{N} \frac{F_1^i}{\lambda_1}}{N_j^k}$$

This duality relationship implies thus that the score of a population unit on the first factor is equal to the average of the standardized factorial weights of the K categories to which it belongs. Conversely the weight of a given category is equal to the average of the standardized scores of the population units belonging to the corresponding category.

[26] Very similar results can be derived for the other factorial axes.

Appendix 3: On Frontier Efficiency Measurement:

1) Duality and the Concept of Output Distance Function in Production Theory:

Let P(x) represent the set of all output vectors y which can be produced using the input vector x. That is P(x) = { y : x can produce y }

The output distance function $d_o(x,y)$ is then defined as

$$d_o(x,y) = \text{Min}\ \{\delta: (y/\delta) \in P(x)\ \}.$$

The following properties of the output distance function may be proven:

(1) $d_o(x,y)$ is increasing in y and decreasing in x;

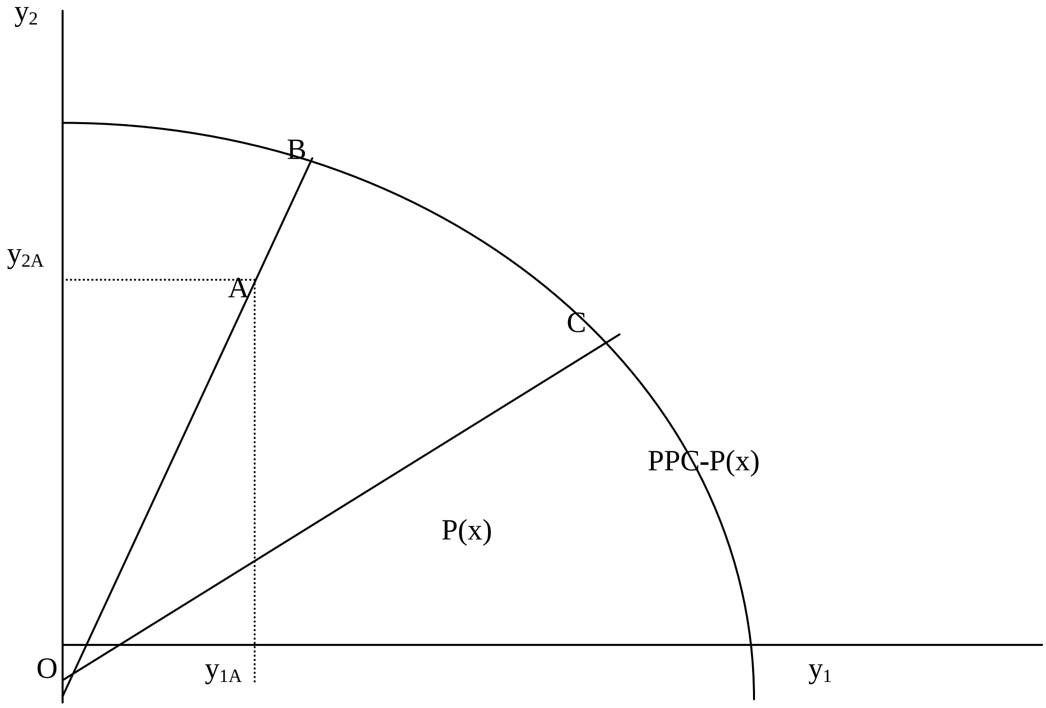
(2) $d_o(x,y)$ is linearly homogeneous in y;

(3) if y belongs to the production possibility set of x (i.e. $y \in P(x)$), then $d_o(x,y) \leq 1$;

(4) the output distance function is equal to unity if y belongs to the "frontier" of the production possibility set (to the production possibility curve of x).

An illustration of this concept is given in Figure 1.

Figure 1



2) Deriving Productivity Indices:

Use will be made of output-oriented indices which focus on the maximal level of outputs that could be produced using a given input vector and a given production technology relative to the observed level of outputs. Their computation is based on the use of output distance functions. The period-s Malmquist productivity index $m_o^s\ (y_s,y_t,x_s,x_t)$ is defined as

$$m_o^s\ (y_s,y_t,x_s,x_t) = d_o^s\ (y_t, x_t) / d_o^s(y_s, x_s)$$

If we assume that the firm is technically efficient in both periods, $d_o^s\ (y_s, x_s) = 1$ so that

$$m_o^s\ (y_s,y_t,x_s,x_t) = d_o^s\ (y_t, x_t)$$

One can define in a similar way an output-oriented Malmquist productivity index based on period-t technology.

3) Estimation Procedures: The Stochastic Production Frontier Approach

Let us take as a simple illustration the case of a Cobb-Douglas production function. Let $\ln y_i$ be the logarithm of the output of firm i (i=1 to N) and x_i a (k+1) row vector, whose first element is equal to one and the others are the logarithms of the k inputs used by the firm. We may then write that

$$\ln (y_i) = x_i\ \beta - u_i \qquad\qquad i = 1 \text{ to } N.$$

where β is a (k+1) column vector of parameters to be estimated and u_i a non-negative random variable, representing the technical inefficiency in production of firm i.

The ratio of the observed output of firm i to its potential output will then give a measure of its technical efficiency TE_i so that

$$TE_i = y_i / \exp (x_i\ \beta) = \exp (x_i\ \beta - u_i) / \exp (x_i\ \beta\ ) = \exp (-u_i)$$

One of the methods allowing the estimation of this output-oriented Farrell measure of technical efficiency TE_i (see, Farrell, 1957) is to use an algorithm proposed by Richmond (1974) which has become known as corrected ordinary least squares (COLS). This method starts by using ordinary least squares to derive the (unbiased) estimators of the slope parameters. Then in a second stage the (negatively biased) OLS estimator of the intercept parameter β_0 is adjusted up by the value of the greatest negative residual so that the new residuals have all become non-negative. Naturally the mean of the observations does not lie any more on the estimated function: the latter has become in fact an upward bound to the observations.

One of the main criticisms of the COLS method is that it ignores the possible influence of measurement errors and other sources of noise. All the deviations from the frontier have been assumed to be a consequence of technical inefficiency. Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) independently suggested an alternative approach called the stochastic production frontier method in which an additional random error v_i is added to the non-negative random variable u_i. We therefore write

$$\ln (y_i) = x_i \beta + v_i - u_i$$

The random error v_i is supposed to take into account factors such as the weather, the luck, etc...and it is assumed that the v_i's are i.i.d. normal random variables with mean zero and constant variance σ_v^2. These v_i's are also assumed to be independent of the u_i's, the latter being taken generally to be i.i.d. exponential or half-normal random variables. In the latter case where the u_i's are assumed to be i.i.d truncations (at zero) of a normal variable $N(0,\sigma)$, Battese and Corra (1977) suggested to proceed as follows. Calling σ_s^2 the sum $\sigma^2 + \sigma_v^2$, they defined the parameter γ as $\gamma=\sigma^2/\sigma_s^2$ (so that γ has a value between zero and one) and showed that the log-likelihood function could be expressed as

$$\ln(L) = -(N/2) \ln(\pi/2) - (N/2) \ln(\sigma_s^2) + \sum_{i=1 \text{ to } N} [1-\Phi(z_i)]-(1/(2\sigma_s^2))\sum_{i=1 \text{ to } N} (\ln y_i-x_i\beta)^2$$

where $z_i =((\ln y_i - x_i \beta)/\sigma_s) \sqrt{(\gamma/(1-\gamma))}$ and $\Phi(.)$ is the distribution function of the standard normal random variable.

The Maximum Likelihood estimates of β, σ_s^2 and γ are obtained by finding the maximum of the log-likelihood function defined previously where this function is estimated for various values of γ between zero and one. More details on this estimation procedure is available in programs such as FRONTIER (Coelli, 1992) or LIMDEP (Green, 1992). The same methods (COLS and Maximum Likelihood) may naturally be also applied when estimating distance functions.

4) The case of multiple inputs and a latent output:

Assume a vector of inputs x , a vector of outputs y and an output distance function $D_{out}(x,y)$ defined as

$$D_{out}(x,u) = \text{Min } \{\theta: (u/\theta) \in P(x)\}$$

where $P(x)$ is the set of all output vectors which can be realised with the input vector x. We can estimate efficiency E via a Bergson-Moorsteen output quantity index

$$E(x, y^s, y^t) = D_{out}(x, y^s)/D_{out}(x, y^t)$$

where y^s and y^t are two output vectors and x is an input vector. Clearly, the further inside the set $P(x)$ an output vector is, the more it must be radially expanded in reach the frontier $P(x)$ and hence the lower the efficiency of the individual.

There is however a problem, that of choosing a reference vector, that is, a specific input vector x. In order to do so we define a N-dimensional vector e of ones, that is, we will assume that each individual is endowed with one unit of each resource. This implies that we define a reference set $P(e)$ which bounds from above the observed output vectors of the various individuals. If an individual has a vector of outputs which places him on the frontier $P(e)$, this implies that she has the maximum level of output and, hence, an output index of unity. Individuals with smaller output levels will have a index values below unity. Note that the index we derive is independent of the units of measurement of the outputs.

To estimate the output distance functions we proceed as follows. We assume a translog functional form

$$\ln(y_M^{-1}) = \beta_0 + \sum_{f=1}^{M-1} \beta_f \ln v_f + \tfrac{1}{2} \sum_{f=1}^{M-1} \sum_{h=1}^{M-1} \beta_{fh} \ln v_f \ln v_h + \varepsilon$$

where $v_f = (y_f / y_M)$, $f = 1,\ldots, M$-1. We may use either COLS or maximum likelihood methods to obtain estimates of the *ß* coefficients. The (modified) residuals which are then derived will provide output distance functions for each individual by means of the transformation

$$D_{out}(e, y^i) = \exp\{\min(\varepsilon_i) - \varepsilon_i\}$$

This distance will by definition be smaller than one (since its logarithm will be negative or at most equal to zero) so that all individual output vectors will lie on or beneath the frontier corresponding to $P(e)$. Hence, the output distance function $D_{out}(e, y^i)$ gives the maximum amount by which individual output vectors must be radially scaled up in order to reach the frontier. Finally, an efficiency $E(x, u^s, u^t)$ is obtained by dividing all the output distance functions by the maximum observed distance —by definition equal to 1.

5) The Case of Multiple Inputs and Outputs:

We estimate the Transformation Efficiency index, TE, using a Malmquist productivity index

$$TE(x^s,x^t,y^s,y^t) = D_{out}(x^s,y^s)/ D_{out}(x^t,y^t).$$

Note that now the reference set $P(x^i)$ is defined as

$$P(x^i) = \{(y^i/\theta): D_{out}(x^i,(y^i/\theta)) = 1\}, \quad i = 1,....,I$$

where I is the number of individuals. All individuals will therefore be compared to the relevant reference set and for those who are able to convert relatively small resources into relatively large outputs the distance function will be unity while less efficient individuals will have a smaller score. The same technique used previously in estimating distance functions is applied here. Note, however, that this time both input and output data are used. The translog output distance function can then be expressed as

$$\ln(y_M^{-1}) = \delta_0 + \sum_{j=1}^{N-1}\alpha_j \ln x_j + \tfrac{1}{2}\sum_{j=1}^{N-1}\sum_{k=1}^{N-1}\alpha_{jk} \ln x_j \ln x_k + \sum_{f=1}^{M-1}\beta_f \ln v_f +$$
$$+ \tfrac{1}{2}\sum_{f=1}^{M-1}\sum_{h=1}^{M-1}\beta_{fh} \ln v_f \ln v_h + \sum_{f=1}^{M-1}\sum_{j=1}^{N-1}\gamma_{fj} \ln v_f \ln x_j + \varepsilon$$

Again, coefficient estimates may be obtained using either COLS or maximum likelihood methods. The (modified) residuals which are then derived provide output distance functions for each individual.

$$D_{out}(x^i,y^i) = \exp\{\max(\varepsilon_i) - \varepsilon_i\}$$

This distance will by definition be smaller than one (since its logarithm will be negative or at most equal to zero) so that all individual input and output vectors will lie on or beneath the frontier $P(x)$.

These output distance functions measure the efficiency with which individuals convert their inputs into outputs. Since, by construction, the maximum observed output distance function is unity, the distance $D_{out}(x^s,y^s)$ will also be equal to the Malmquist Productivity Index $TE(x^s,x^t,y^s,y^t)$ — when divided by the maximum output distance.

Appendix 4: On the Concept of Shapley Decomposition

The concept of Shapley (1953) decomposition is a technique borrowed from game theory but extended to applied economics by Shorrocks (1999) and Sastre and Trannoy (2002). Let us explain it briefly.

Assume an indicator I is a function of three determinants a,b,c and is written as $I = I(a,b,c)$. I could be an index of inequality, the R-square of a regression and more generally any function of variables, this function being linear or not.

There are obviously 3!=6 ways of ordering these three determinants a, b and c:

$$(a,b,c),(a,c,b),(b,a,c),(b,c,a),(c,a,b),(c,b,a) \qquad (4\text{-}1)$$

Each of these three determinants may be eliminated first, second or third. The respective (marginal) contributions of the determinants a,b,c will hence be a function of all the possible ways in which each of these determinants may be eliminated. Let for example $C(a)$ be the marginal contribution of a to the indicator $I(a,b,c)$.

If a is eliminated first its contribution to the overall value of the indicator I will be expressed as $I(a,b,c) – I(b,c)$ where $I(b,c)$ corresponds to the case where a is equal to zero. Since expression (4-1) indicates that there are two cases in which a appears first and may thus be eliminated first, we will give a weight of (2/6) to this possibility.

If a is eliminated second, it implies that another determinant has been eliminated first (and been assumed to be equal to 0). Expression (4-1) indicates that there are two cases in which this possibility occurs, the one denoted in (4-1) as (b,a,c) and the one denoted (c,a,b). In the first case the contribution of a will be written as $I(a,c) – I(c)$ while in the second it is expressed as $I(a,b) – I(b)$. To each of these two cases we evidently give a weight of (1/6).

Finally if a is eliminated third, it implies that both b and c are assumed to be equal to 0. Expression (4-1) indicates that there are two such cases, the one denoted (b,c,a) and the one denoted (c,b,a). Since we may assume that when each of the three determinants is equal to 0, the indicator I is equal to 0, we may write that the contribution of a in this case will be equal to $I(a) – 0 = I(a)$ and evidently we have to give a weight of (2/6) to such a possibility since there are two such cases.

We may therefore summarize what we have just explained by stating that the marginal contribution $C(a)$ of the determinant a to the overall value of the indicator I may be written as

$$C(a) = (2/6)[I(a,b,c) – I(b,c)] + (1/6)[I(a,c) – I(c)] + (1/6)[I(a,b) – I(b)] + (2/6)I(a) \qquad (4\text{-}2)$$

One can similarly determine the marginal contribution C(b) of b and C(c) of c and then find out that

$I(a,b,c) = C(a) + C(b) + C(c)$	(3-3)

Appendix 5: The determinants of efficiency in the five countries analyzed

Table 5-1: Estimating for Brazil the impact of non discretionary inputs on the efficiency of transforming discretionary inputs into outputs (test scores)

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.65073	0.09210		
Explanatory Variables:				
Constant			0.67755	67.75
Gender of the student	0.53299	0.49891	-0.00321	-0.49
Human capital of the parents	0.01208	0.53890	0.00373	0.51
Material wealth of the parents	0.09923	0.84353	0.00263	0.54
Information on school governance: School funding	-0.26993	1.29254	-0.00975	-3.10
Information on school governance: Autonomy of school	-0.03478	0.43228	0.00205	0.24
Information on school governance: Transparency of information	-0.01508	0.51300	0.00016	0.03
Information on school governance: Homogeneity of school	-0.06433	0.84695	0.00137	0.35
Location of school: village, hamlet or rural area	0.01523	0.12246	-0.08727	-3.07

Location of school: small town (from 3,000 to 15,000 people)	0.17386	0.37899	-0.03081	-2.50
Location of school: town (15,000 to 100,000 people)	0.38452	0.48648	-0.02798	-2.64
Location of school: city (100,000 to 1000,000 people)	0.29315	0.45521	-0.03542	-3.26
Importance of learning efforts in the eyes of the student	-0.00008	0.70921	-0.00498	-1.07
Self-rated ability of student	0.00561	0.77078	-0.01114	-2.62

Note: R-square: 0.05746; Adjusted R-Square: 0.04163; Number of observations: 788

Table 5-2: Estimating for Chile the impact of non discretionary inputs on the efficiency of transforming discretionary inputs into outputs (test scores)

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.63257	0.08438		
Explanatory Variables:				
Constant			0.64279	107.08
Gender of the student	0.40909	0.49167	-0.00938	-1.53
Human capital of the parents	-0.00852	0.56215	0.02986	4.83
Material wealth of the parents	0.10628	0.76676	-0.01231	-2.69
Information on school governance: **School funding**	0.04287	1.07248	0.00357	1.13
Information on school governance: **Autonomy of school**	-0.14497	0.52446	0.03311	5.37
Information on school governance: **Transparency of information**	-0.06671	0.68359	0.00052	0.10
Information on school governance: **Homogeneity of school**	-0.13792	1.13439	0.00479	1.54
Location of school: village, hamlet or rural area	0.01136	0.10599	-0.07171	-2.51

Location of school: small town (from 3,000 to 15,000 people)	0.12642	0.33232	-0.01333	-1.07
Location of school: town (15,000 to 100,000 people)	0.23438	0.42361	0.00748	0.87
Location of school: city (100,000 to 1000,000 people)	0.32102	0.46687	0.00467	0.60
Importance of learning efforts in the eyes of the student	-0.00143	0.69902	-0.01020	-2.37
Self-rated ability of student	0.01834	0.80287	-0.01312	-3.48

Note: R-square: 0.16623; Adjusted R-Square: 0.15052; Number of observations: 704

Table 5-3: Estimating for Colombia the impact of a limited set of non discretionary inputs on the efficiency of transforming discretionary inputs into test scores
(the explanatory variables are identical to those used for the four other Latin American countries)

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.67458	0.09384		
Explanatory Variables:				
Constant			0.70990	91.90
Gender of the student	0.55331	0.49715	-0.04068	-6.68
Human capital of the parents	0.00904	0.53264	-0.02567	-4.04
Material wealth of parents	0.06703	0.70412	-0.00713	-1.44
Information on school governance: School funding	-0.13293	0.78579	-0.00437	-0.98
Information on school governance: Autonomy of school	0.24047	0.58610	-0.00909	-1.54
Information on school governance: Transparency of information	0.01173	0.50055	0.01517	2.29
Information on school governance: Homogeneity of school	0.01320	0.81932	0.00633	1.69
Location of school: village, hamlet or rural area	0.06734	0.25061	-0.02403	-1.78

Location of school: small town (from 3,000 to 15,000 people)	0.21886	0.41347	-0.03765	-3.75
Location of school: town (15,000 to 100,000 people)	0.22110	0.41499	-0.00862	-0.96
Location of school: city (100,000 to 1000,000 people)	0.22222	0.41574	0.00455	0.51
Importance of learning in eyes of student	-0.00022	0.69964	-0.00863	-1.98
Self-rated ability of student	0.00060	0.79228	-0.01057	-2.73

Note: R-square: 0.12795; Adjusted R-Square: 0.11502; Number of observations: 891

Table 5-4: Estimating for Mexico the impact of a limited set of non discretionary inputs on the efficiency of transforming discretionary inputs into test scores

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.68198	0.08631		
Explanatory Variables:				
Constant			0.68565	166.97
Gender of the student	0.55025	0.49747	-0.01721	-6.31
Human capital of the parents	0.00975	0.54609	-0.02705	-9.56
Material wealth of parents	-0.15599	0.89566	0.00288	1.61
Information on school governance: **School funding**	0.12513	0.89822	-0.00615	-3.69
Information on school governance: **Autonomy of school**	0.05031	0.51317	-0.00416	-1.47
Information on school governance: **Transparency of information**	-0.02640	0.54355	0.01123	4.34
Information on school governance: **Homogeneity of school**	0.05264	0.94493	0.00641	4.37
Location of school: village, hamlet or rural area	0.06384	0.24446	-0.02432	-3.56

Location of school: small town (from 3,000 to 15,000 people)	0.17990	0.38410	-0.00217	-0.43
Location of school: town (15,000 to 100,000 people)	0.21366	0.40989	0.00556	1.16
Location of school: city (100,000 to 1000,000 people)	0.40148	0.49020	0.02052	4.83
Importance of learning in eyes of student	0.00049	0.74687	-0.00537	-2.90
Self-rated ability of student	0.00552	0.78876	-0.00573	-3.27

Note: R-square: 0.08079; Adjusted R-Square: 0.07763; Number of observations: 3791

Table 5-5: Estimating for Uruguay the impact of a limited set of non discretionary inputs on the efficiency of transforming discretionary inputs into test scores

Dependent Variable: Individual standardized efficiency scores $d(k)$

Variables	Mean of the variable	Standard Deviation of the variable	Coefficient of the variable in the regression	t-value of the variable
Dependent Variable	0.66854	0.09848		
Explanatory Variables:				
Constant			0.67153	105.42
Gender of the student	0.53711	0.49862	-0.00237	-0.41
Human capital of the parents	-0.00579	0.56534	0.04413	7.16
Material welath of parents	0.08020	0.76287	-0.01244	-2.78
Information on school governance: **School funding**	-0.37233	1.17358	-0.00391	-0.95
Information on school governance: **Autonomy of school**	0.17556	0.60110	-0.01944	-2.85
Information on school governance: **Transparency of information**	0.00308	0.87397	0.00073	0.21
Information on school governance: **Homogeneity of school**	0.02829	0.98752	-0.00416	-1.39
Location of school: village, hamlet or rural area	0.04980	0.21754	0.00254	0.18

Location of school: small town (from 3,000 to 15,000 people)	0.14551	0.35261	0.00283	0.28
Location of school: town (15,000 to 100,000 people)	0.29883	0.45774	0.00441	0.59
Location of school: city (100,000 to 1000,000 people)	0.05078	0.21955	-0.00011	-0.01
Importance of learning in eyes of student	-0.00221	0.73269	0.01168	2.90
Self-rated ability of student	0.01071	0.82274	-0.01968	-5.48

Note: R-square: 0.14183; Adjusted R-Square: 0.13079; Number of observations: 1024