The efficiency of education in generating literacy: a stochastic frontier approach

de Witte, K.; Groot, W.N.J.; Maassen van den Brink, H.

Published in:
Review of Economics & Finance

Review of Economics & Finance
Submitted on 18/October/2011
Article ID: 1923-7529-2012-01-25-13
Kristof De Witte, Dr. Wim Groot and Henriëtte Maassen van den Brink

The Efficiency of Education in Generating Literacy: A Stochastic Frontier Approach

Dr. Kristof De Witte[1, 2]

1. University of Maastricht and University of Amsterdam
Top Institute for Evidence Based Education Research
Postbus 616, 6200 MD Maastricht, The NETHERLANDS
Tel: +31(0)43 38 84 437 E-mail: k.dewitte@maastrichtuniversity.nl
2. Faculty of Economics and Business, University of Leuven (KUL)
Naamsestraat 69, 3000 Leuven, BELGIUM
E-mail: Kristof.dewitte@econ.kuleuven.be

Prof. Dr. Wim Groot, Prof. Dr. Henriëtte Maassen van den Brink

University of Maastricht and University of Amsterdam
Top Institute for Evidence Based Education Research
Postbus 616, 6200 MD Maastricht, The NETHERLANDS
E-mail of Prof. Dr. Wim Groot: Wim.groot@maastrichtuniversity.nl
Prof. Dr. Henriëtte Maassen van den Brink: h.maassenvandenbrink@uva.nl

Abstract: The growing importance attached to education as a key factor to improve economic performance coupled with the persistent scarcity of resources for education makes it important that skills and literacy are produced efficiently. This paper provides an international comparison of the efficiency of literacy production. We find substantial differences between countries in levels of literacy, differences in literacy between education levels and differences in the efficiency of literacy production. The findings suggest that in almost all countries the scope for efficiency improvements in education is large. So even without major increases in (public) funding, improvements in educational outcomes are achievable. We can get better value for the money we spend on education.

JEL Classifications: I22, C14
Keywords: Literacy, Education, Stochastic frontier analysis, Meta-frontier

1. Introduction

Expenditures on education differ between countries. According to OECD (2006), in 2002 Denmark, Iceland, Korea and the United States had the highest spending on education (more than 7% of GDP). Eight out of 28 OECD countries spend less than 5% on education. Although there is some heterogeneity among countries in the educational budget, in all countries the budget is increasing (in both nominal and real terms). According to OECD (2006), during 1995 and 2002 in OECD countries, average real total expenditures on education increased by 5%. This raises the question whether the money for education is spend effective (i.e., doing the right things) and efficient (i.e., doing the things right). Little is known about the outputs that are obtained by the educational investments. This question is especially relevant, given the importance attached to education. Human capital acquired through education is thought by many people as the single most important source to attain a competitive advantage in a globalizing economy (see Becker, 1975; Schultz, 1967 and many references thereto). Education is generally thought to foster economic growth and to contribute to quality of life and the development of values that generate social cohesion and an open society (Willms, 1999).

In the paper at hand, we focus on one component of education: literacy. In doing so, we follow the interpretation of Unesco in defining literacy as '*... the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning to enable an individual to achieve his or her goals, to develop his or her knowledge and potential, and to participate fully in the wider society*'. We focus on literacy as literacy is considered to be an important indication (if not the most important indication) for the distribution of the economic and cultural capital (cf. Willms, 1999, 1997). Literacy is captured from the 1994-1996 International Adult Literacy Survey (IALS), which is an OECD survey on issues as literacy, schooling, and employment. The survey has been extensively used in estimating the returns of education (Blau & Kahn, 2005 and references therein).

This paper contributes to the literature in three aspects. Firstly, it examines how well education systems are able to produce literacy among their population, i.e. it looks at the efficiency of literacy production in a number of OECD countries. In particular, we apply stochastic frontier techniques (Aigner *et al.*, 1977; Meeusen & Van Den Broeck, 1977) on literacy scores to evaluate the impact of education and social background on literacy scores. As such, we investigate the effectiveness of schooling across countries: to which extent does an additional year of schooling increases literacy and does this differ across countries? In doing so, we account for various background characteristics which capture the socio-economic status of the respondent.

Secondly, we estimate the impact of socio-economic variables on the generation of efficiency in a cross-country perspective. By comparing the coefficient estimates of the father and mother of the respondent to the respondent's literacy, we deduce insights on the strength of role models across countries. We analyse whether the education of the mother or of the father has the strongest impact on the respondent's literacy, and whether there is some explanation for this effect.

Finally, using the metafrontier framework of Battese & Rao (2002) and O'Donnell *et al.* (2008), we estimate the inefficiency within and across countries. The metafrontier framework makes a distinction between country specific inefficiency and inefficiency of the educational system. As such, we estimate the scope for efficiency improvements in education and detect best-practice countries (i.e., countries that are able to obtain higher literacy outcomes with a given budget per pupil).

The paper unfolds as follows. Section 2 provides a brief review of previous work and highlights the differences with the paper at hand. In Section 3, we outline the traditional Stochastic Frontier Analysis model. A fourth section describes the data and provides some preliminary insights. Section 5 discusses the estimated SFA coefficients and estimates the relative (in)efficiency across countries. We finally conclude the paper.

2. A Brief Literature Review

This paper relates to a wide range of academic literature on the performance estimations in education. This section does not intend to provide an exhaustive overview of the literature, on the contrary. It intends to clarify our contributions on three aspects to this literature by pointing to some influential references.

Firstly, consider the adapted methodology of stochastic frontier estimation (see next section). A few studies have applied stochastic frontier estimation techniques to evaluate school efficiency. Izadi *et al.* (2002) apply stochastic frontier estimation to analyze inefficiencies in higher education in Great Britain. They conclude that inefficiencies in higher education are fairly modest and on the margin of statistical significance. Barbetta & Turati (2003) use stochastic frontier analysis to evaluate the efficiency of junior high schools in Italy. They conclude that the proprietary structure affects efficiency: not for profit schools are more efficient than public ones, whereas for profit

schools are less efficient than public schools. Furthermore, foreign and disabled students affect efficiency of schools negatively. Ruggiero & Vitaliano (1999) use stochastic frontier techniques at a more aggregate level, i.e. to examine the efficiency of New York school districts. They find that mean inefficiency of school districts amounts to 14%, which indicates a significant scope for improvement.

However, a potential limitation of stochastic frontier analysis is that the residual is interpreted as representing technical inefficiency but that it may also represent other systematic and unobservable differences between units of analysis (schools, countries, firms, etc.). Newhouse (1994) criticises the approach because of its difficulty to disentangle inefficiency from quality differences: greater technical inefficiency may simply represent better quality. This problem occurs, e.g., when school expenditures or the number of graduates are taken as an output measure in a stochastic frontier cost function. Newhouse (1994) argues that stochastic frontier techniques can best be applied when the outcome measure is homogeneous and uni-dimensional. However, even years of education – a homogenous and uni-dimensional educational outcome measure – poses difficulties in disentangling efficiency from quality as not all years of education are of similar quality. We believe that production of literacy is a more homogeneous and uni-dimensional outcome measure than school expenditures or the number of graduates and that this solves some of the criticism on the stochastic frontier technique.

Secondly, consider the cross-national focus of the paper. Other studies have used comparative data to evaluate differences in school performance across countries. Hanushek & Luque (2003) use data from the Third International Mathematics and Science Study (TIMSS) to compare performance of schooling systems of more than 40 countries. They conclude that the impact of schooling resources on student performance is rather limited. They further conclude that organizational characteristics and incentives may be more important in explaining performance differences than the level of resources allocated to schools. Finally they find that the impact of school resources does not vary systematically with country income or development. The Hanushek & Luque (2003) study differs from ours in that we compare the performance of schooling systems at a more aggregate level. We also use data from population wide samples, rather than the pupil-based samples of the TIMSS, and use a more general outcome measure – literacy - than mathematics and science test scores. A similar focus on literacy can be found in Charette & Meng (1997) who explored the determinants of literacy using the Canadian 'Literacy Skills used in Daily Activities' (LSUDA). Our paper differs from the latter as we consider a cross-country perspective in a different data set (IALS).

Finally, consider our data set. We use the data of the International Adult Literacy Survey (IALS) to estimate a stochastic frontier model to analyze the performance of schooling systems in 11 countries. The IALS data have – among others - been used by Blau & Kahn (2005) to analyze the role of cognitive skills on wage inequality in the USA. This study finds that the greater dispersion of cognitive skills among people in the USA partly explains the higher wage inequality. Denny *et al.* (2000) use the IALS data to examine the impact of functional literacy on earnings. They conclude that literacy has an effect on earnings but that education remains the dominant factor. Our paper contrasts to previous literature using IALS in that we focus on the efficiency of literacy provision (at a cross-national level). An issue not explored before.

3. A Stochastic Frontier Model of Literacy

In the paper at hand, we presume that the educational system focuses on educating its citizens in terms of literacy. Assume there exists a production function which produces literacy (L) by combining educational inputs (S), social or parental environment inputs (P) and other characteristics (X). Following standard economic theory, the production function is presented as:

$$L = f(S, P, X) + \varepsilon \qquad (1)$$

where f(.) denotes the production technology and ε an error term. Instead of considering a normal distributed ε, Aigner *et al.* (1977) and Meeusen & Van Den Broeck (1977) decomposed ε into two terms: $\varepsilon = \nu - \mu$. Firstly, ν denotes a random component which captures the stochastic elements and unobserved heterogeneity. ν is assumed to be distributed along $N(0, \sigma_\nu)$. Secondly, μ represents an inefficiency term and captures the shortfall in output given the inputs. As efficiency should be positive, μ is assumed to be half-normal distributed: $\mu \sim \left| N(0, \sigma_\mu) \right|$ (see Kumbhakar & Lovell (2003) for an extensive discussion).

Although the parametrization of a learning process is quite problematic (see, e.g., Hanushek, 1979, 1986), we assume a simple linear specification for the literacy production function:

$$L = \beta_0 + \beta_1 S + \beta_2 P + \beta_3 X + \nu - \mu \qquad (2)$$

where β are vectors of coefficients that measure the impact of the inputs on literacy. The model is estimated by maximum likelihood for which the loglikelihood of the model is given by:

$$LogL = -Log(\sigma) + \sum Log(1 - \Phi(-\frac{\varepsilon\lambda}{\sigma})) - \sum \frac{\varepsilon^2}{2\sigma^2} \qquad (3)$$

where $\sigma = \sigma_\mu + \sigma_\nu$ and $\lambda = \sigma_\mu / \sigma_\nu$ and Φ(.) is the standard normal distribution function.

The estimation results can be used to obtain a measure of technical inefficiency. Jondrow *et al.* (1982) derive an estimate of technical inefficiency, i.e. the conditional expectation of μ given ε:

$$E(\mu \mid \varepsilon) = \frac{\varepsilon\lambda / \sigma + \varphi(-\varepsilon\lambda / \sigma)}{(1 - \Phi(-\varepsilon\lambda / \sigma))\sigma_\mu \sigma_\nu / \sigma} \qquad (4)$$

We estimated the stochastic frontier in Stata 11. To avoid infeasibility in the maximum likelihood estimation due to wrongly specified starting values, which is common in samples with low inefficiency, we specified starting values by a simple linear regression (see Kumbhakar & Lovell, 2003).

4. Data and Descriptive Analysis

To examine the relationship between the education system and the level of literacy, we obtain data from the OECD 'International Adult Literacy Survey' (IALS). The IALS survey contains micro data for 38,358 persons in 11 OECD countries: Belgium, Canada, Germany, Ireland, Netherlands, New Zealand, Poland, Sweden, Switzerland, the United Kingdom and the United States.[1] For each of the 11 countries, the survey contains a sample of persons aged over 16, who were similarly tested and interviewed on their literacy. To make the sample representative, the data set includes weights for each individual. In the remainder of the paper, we apply these weights.

The IALS aims at investigating how well adults use information to function in society. Additionally, it allows us to determine the factors that influence literacy proficiency and to compare these across countries.[2] Literacy (L; the dependant variable) is captured by five plausible values (ranging from 0 to 500) on three complementary dimensions: prose (i.e., understand information from texts), document (i.e., use information from forms, schedules, maps or charts) and quantitative (i.e., apply arithmetic operations) literacy. By using the three dimensions and the scale, the IALS

[1] After removing respondents with incomplete information on relevant variables, we obtain a data set of 31,984 observations.

[2] Although some authors still doubt the comparability across countries (e.g., Blum *et al.*, 2001), we follow the OECD in assuming that the literacy scores are comparable across countries.

intended to construct a comprehensive measure for illiteracy and avoid the simplistic notion that people are either literate or illiterate. Following previous literature, we take the arithmetic average of the three dimensions on each of their 5 plausible values (note that the first principle component resulted in comparable scores). As such, an aggregated variable capturing literacy is obtained.[3]

Following the OECD (2000) the obtained score indicates literacy. A score between 0-225 identifies people who have very poor literate skills. On the three aggregated dimensions, i.e. our literacy score, about 19.80% of the sample turns out to be very poorly literate. Scores between 226 and 275 denote respondents with weak literacy skills (27.12% of the sample). A third level (with values between 276 and 325) is reserved for people with a literacy level which suffices for dealing with a complex society (37.46% of the sample). While a fourth and fifth level (respectively, 326-375 and 376-500) indicates respondents who can cope with higher-order information processing (14.56% and 1.06% of the sample).

Figure 1 presents the heterogeneity in literacy (L) among the 11 countries. There are clearly some differences in both the (weighted) mean literacy, and the spread of literacy (in terms of standard deviation and range) across the countries. Swedish and Dutch citizens have the highest average literacy, while Polish citizens have the lowest. The spread of literacy scores is most equal in Germany and the Netherlands, while the inequality of literacy is highest in the French speaking Canada, Northern Ireland and Poland (we will test this in the next section more formally by using SFA estimates). The large variation between the maximum and minimum literacy score can also be observed from the summary statistics in Table 1. It should be noted in this respect that the sample includes for each country a representative sample of people older than 16 years. For example, the correlation between age and schooling of the mother is with -0.187 large and significant. The nonparametric spearman rank correlation is even higher with a correlation of -0.26.

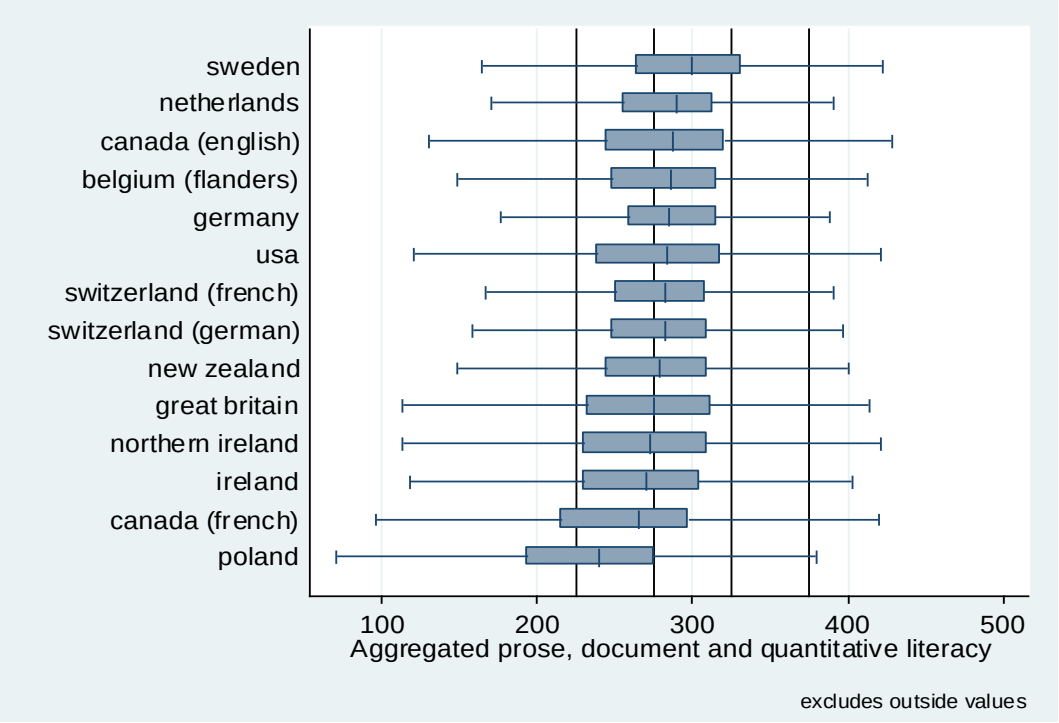

Figure 1 Distribution of scores across the countries

[3] Note that an alternative approach could consist of including all three dimensions as outputs and estimating a distance function rather than a production function applied here. We consider this as scope for further research.

Table 1 Descriptive statistics

Canada - English	Obs.	Mean	St. Dev.	Min.	Max.	Germany	Mean	St. Dev.	Min.	Max.
Literacy	3951	276.330	67.241	54.625	443.508	1832	285.870	42.264	102.620	418.570
Own school	3951	4.072	1.459	1	9	1832	2.842	1.392	1	10
Schooling mother	3951	3.991	2.104	1	9	1602	2.870	1.408	1	10
Schooling father	3951	4.115	2.289	1	9	1602	2.189	0.777	1	6
Gender (1 = female)	3951	1.512	0.500	1	2	1602	2.438	1.165	1	7
Age	3951	41.598	16.273	16	70	1832	40.009	13.670	16	65
Migrant (1=yes)	3951	1.258	0.438	1	2	1832	1.060	0.237	1	2
Canada - French						USA				
Literacy	1709	255.192	65.600	13.052	419.098	2905	274.897	64.022	40.690	438.109
Own school	1709	3.546	1.559	1	9	2905	3.765	1.712	1	7
Schooling mother	1709	3.068	2.378	1	9	2434	3.940	1.704	1	7
Schooling father	1709	3.099	2.409	1	9	2394	3.014	1.579	1	7
Gender	1709	1.516	0.500	1	2	2321	3.186	1.869	1	7
Age	1709	42.495	16.182	16	70	2905	39.676	13.391	15	89
Migrant (1=yes)	1709	1.087	0.282	1	2	2905	1.126	0.332	1	2
Switzerland - German						Ireland				
Literacy	1244	276.581	51.769	89.634	403.601	2362	262.936	57.039	53.575	402.277
Own school	1171	3.340	1.439	1	9	2304	2.780	1.761	0	9
Schooling mother	1171	2.587	1.489	1	10	2133	2.832	1.773	0	9
Schooling father	1171	3.252	1.753	1	10	2084	1.671	1.364	0	10
Gender	1244	1.500	0.500	1	2	2051	1.699	1.552	0	10
Age	1244	39.782	13.395	16	65	2362	36.614	13.524	16	64
Migrant (1=yes)	1244	1.156	0.363	1	2	2362	1.059	0.235	1	2
Switzerland - French						Netherlands				
Literacy	1328	277.802	49.100	56.591	390.872	3090	281.360	46.915	56.627	417.181
Own school	1255	3.575	1.882	1	9	3090	2.974	1.792	0	10
Schooling mother	1255	2.514	1.956	1	10	2816	3.015	1.776	0	10
Schooling father	1255	3.130	2.147	1	10	2792	1.783	1.340	0	9
Gender	1328	1.505	0.500	1	2	2779	2.323	1.709	0	9
Age	1328	39.091	13.716	16	65	3090	41.143	15.620	16	74
Migrant (1=yes)	1328	1.247	0.431	1	2	3090	1.065	0.247	1	2
United Kingdom						Poland				
Literacy	3731	269.154	60.722	19.712	413.454	2999	229.463	64.088	25.103	379.313
Own school	3329	2.9076	1.544	0.000	7	2816	2.516	1.532	0	9
Schooling mother	3101	2.3896	1.461	0.000	10	2789	1.681	1.878	0	10
Schooling father	2971	2.5271	1.619	0.000	10	2777	1.801	1.812	0	10
Gender	3731	1.495	0.500	1	2	2999	1.504	0.500	1	2
Age	3731	39.137	13.757	16	65	2999	37.874	13.539	16	64
Migrant (1=yes)	3731	1.060	0.238	1	2	2999	1.018	0.132	1	2
Northern Ireland						Sweden				
Literacy	2904	265.049	61.959	39.549	469.733	3014	293.338	55.256	50.949	421.872
Own school	2615	2.670	1.449	0	7	2820	3.196	1.877	0	9
Schooling mother	2421	1.978	1.446	0	10	2820	1.979	2.026	0	10
Schooling father	2365	1.987	1.539	0	10	2820	2.173	2.133	0	10
Gender	2904	1.504	0.500	1	2	3014	1.514	0.500	1	2
Age	2904	37.860	13.907	16	65	3014	45.832	18.690	15	94
Migrant (1=yes)	2904	1.039	0.194	1	2	3014	1.089	0.285	1	2
Belgium						New Zealand				
Literacy	2215	277.174	55.075	28.499	413.053	3300	272.010	56.533	43.313	411.833
Own school	1895	2.983	1.581	0	9	2526	3.385	1.679	0	9
Schooling mother	1895	2.465	2.412	0	10	2490	2.783	1.870	0	10
Schooling father	1895	2.714	2.381	0	10	2453	2.994	2.042	0	10
Gender	2215	1.561	0.775	1	9	3300	1.528	0.499	1	2
Age	2215	38.424	13.316	16	64	3300	37.653	13.268	16	65
Migrant (1=yes)	2215	1.032	0.177	1	2	3300	1.196	0.397	1	2

As can be observed from Figure 1 above, the distribution of the literacy scores shows some remarkable differences between countries. In Poland, the country with the lowest literacy, about 41.43% (33.50%) of the respondents is categorized in level 1 (respectively, level 2). On the contrary, in Sweden only 8.33% (19.59%) of the adults have poor and weak literacy skills (level 1 and 2) (see OECD, 2000 for an extensive discussion).

To obtain the estimate of the influence of education on literacy as accurate as possible, we control in the SFA model for various background characteristics. First, we include the respondents' education level (S) as independent variable. Intuitively, a higher respondents' education level will result in a higher literacy score. Education level is made comparable across countries by using the International Standard Classification of Education (ISCED) code,[4] where Isced 0 = pre-primary education; Isced 1 = primary education, Isced 2 = lower secondary education; Isced 3 = upper secondary education; Isced 5 = first stage of tertiary education and Isced 6 = second stage of tertiary education. In the analysis below, we take respondents without with pre-primary as a reference group. For countries where the reference group was too small, the reference group consists of Isced 0 and 1. The (weighted) summary statistics are provided in Table 1.

Second, we control for parental schooling (P): education level of the mother and father. It is commonly acknowledged that higher educated parents raise higher educated children. From Table 1 we observe that, on average, fathers are highest educated in Canada, while Switzerland and the Netherlands have, on average, the lowest educated fathers. Mothers of the respondents obtained the highest education levels in Canada and USA and the lowest levels are observed in Poland.

Third, we control for some exogenous characteristics (X) such as gender (female = 1), age, and migrant status. Some (weighted) summary statistics are presented in Table 1. Although both gender, age and migrant status do not significantly differ across countries, it might have an influence on the individual's literacy score. Intuitively, we expect age and migrant status to be negatively correlated to literacy. A priori, the influence of gender is ambiguous: on average, males underperform females; but females have historically a lower education level.[5]

5. Results of the Stochastic Frontier Estimations

5.1 Explaining Literacy

The results of the SFA estimation are presented in Table 2 on next page. Before we discuss the coefficients, some comments on the interpretation of the results are in order. Firstly, we do not claim to be able to identify the causal effect between education level and literacy. To do so, we need instrumental variables which are not available in the data set at hand. Nevertheless, as we have an adult sample where educational attainment precedes and predetermines literacy achievement in time, there is at least some justification for assuming that education (pre-) determines literacy (and not the other way around).[6] Secondly, to assess the effectiveness of the educational system, we look at the relative differences in literacy between levels of education in a cross-country perspective. In doing so, we control for various other individual characteristics (see Table 1 above).

[4] A similar construction of education level consists of both respondent's education as well as education of mother and father.

[5] In the SFA regressions below, we made robustness tests with gender, age and gender*age. This delivered insignificant results on the latter multiplicative variable.

[6] Note, however, that the achieved education level might be influenced by omitted variables such as the cognitive abilities, which are likely correlated with the literacy scores. Nevertheless, (a proxy for) ability is unfortunately not available in the IALS sample at hand.

The estimation results of the stochastic frontier specification are presented in Table 2. We firstly estimate the model for each of the countries in our sample separately (i.e., relative to the group frontier). We will relax this assumption later.

Table 2 Stochastic frontier estimates of literacy scores (standard errors below the estimates)

	Canada (English)	Canada (French)	Switzerland (German)	Switzerland (French)	Germany	USA	Ireland	Netherlands	Poland	Sweden	New Zealand	United Kingdom	Northern Ireland	Belgium (Flanders)
Intercept	5.77232	5.68233	5.70115	5.78565	5.89814	5.68198	5.17351	6.11960	5.43330	5.52182	5.19517	5.61733	4.55531	6.06179
	-0.00056	-0.00113	-0.00149	-0.00198	-0.00064	-0.00020	-0.00288	-0.00065	-0.00079	-0.00112	-0.00830	-0.00103	-0.01498	-0.00118
Education respondent (Ref. group indicated in the table)														
Isced 0	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	-0.15006	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
								0.00049						
Isced 1	Ref.	Ref.	Ref.	Ref.	0.04940	Ref.	0.31263	-0.12043	0.22044	0.36098	0.17733	0.34625	1.12633	0.09938
					0.00061		0.00104	0.00041	0.00032	0.00072	0.00814	0.00064	0.01395	0.00063
Isced 2	0.18870	0.26088	0.02367	0.00819	0.10702	0.08478	0.42080	-0.03110	0.31785	0.45158	0.40126	0.43398	1.19679	0.16866
	0.00026	0.00025	0.00055	0.00069	0.00057	0.00008	0.00108	0.00041	0.00033	0.00074	0.00805	0.00063	0.01394	0.00063
Isced 3	0.37257	0.35377	0.10156	0.10137	0.14738	0.17561	0.49712	0.02934	0.41054	0.47506	0.49647	0.49669	1.26370	0.20771
	0.00025	0.00025	0.00053	0.00064	0.00057	0.00006	0.00110	0.00041	0.00033	0.00073	0.00805	0.00064	0.01395	0.00064
Isced 5	0.43133	0.46932	0.11790	0.13964	0.19596	0.22003	0.53183	Ref.	0.43957	0.52728	0.52061	0.52849	1.30009	0.26717
	0.00027	0.00028	0.00059	0.00080	0.00058	0.00007	0.00115		0.00036	0.00074	0.00805	0.00064	0.01396	0.00068
Isced 6	0.54086	0.53437	0.17665	0.18067	0.23101	0.29300	0.56824	0.06783	0.48506	0.54105	0.55556	0.56960	1.33541	0.30016
	0.00027	0.00037	0.00061	0.00073	0.00058	0.00007	0.00115	0.00041	0.00036	0.00074	0.00805	0.00064	0.01395	0.00070
Education mother (Ref. group indicated in the table)														
Isced 0	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	0.05846	Ref.	0.07875	0.07448	0.04268	0.03895	0.19336	0.11320
							0.00243		0.00028	0.00059	0.00166	0.00069	0.00454	0.00072
Isced 1	Ref.	Ref.	0.35580	-0.00598	Ref.	Ref.	0.04861	0.04864	0.12984	0.11101	0.09974	0.04912	0.21212	0.10245
			0.00132	0.00104			0.00246	0.00021	0.00029	0.00061	0.00129	0.00060	0.00412	0.00058
Isced 2	0.06713	0.02750	0.38717	-0.01257	0.05511	-0.03378	0.04704	0.05459	0.11730	0.10964	0.10982	0.03755	0.22390	0.09682
	0.00019	0.00021	0.00134	0.00115	0.00019	0.00005	0.00248	0.00022	0.00032	0.00064	0.00128	0.00059	0.00416	0.00059
Isced 3	0.08159	0.00099	0.39834	-0.00543	0.04072	0.03187	0.07778	0.04063	0.12524	0.11208	0.10229	0.04522	0.27059	0.10013
	0.00018	0.00024	0.00134	0.00113	0.00021	0.00004	0.00250	0.00024	0.00033	0.00062	0.00131	0.00061	0.00425	0.00062
Isced 5	0.09017	0.03183	0.42006	-0.06283	0.06321	0.02497	0.01394	Ref.	0.14744	0.12613	0.10167	0.01970	0.28850	0.12115
	0.00023	0.00038	0.00155	0.00144	0.00024	0.00007	0.00260		0.00049	0.00066	0.00133	0.00061	0.00426	0.00068
Isced 6	0.03372	-0.09150	0.45508	-0.03256	0.07993	0.04294	0.06576	0.06492	0.14207	0.11780	0.12392	0.02070	0.22340	0.11865
	0.00018	0.00026	0.00143	0.00134	0.00024	0.00006	0.00264	0.00028	0.00042	0.00067	0.00141	0.00061	0.00430	0.00086
Education father (Ref. group indicated in the table)														
Isced 0	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	0.14943	Ref.	0.16497	0.08698	0.02381	0.06669	0.03208	-0.13491
							0.00280		0.00031	0.00068	0.00164	0.00072	0.00349	0.00080
Isced 1	Ref.	Ref.	0.01494	0.13868	Ref.	Ref.	0.20710	0.01982	0.15280	0.10350	0.05966	-0.07007	0.01985	-0.13211
			0.00158	0.00123			0.00281	0.00026	0.00031	0.00070	0.00148	0.00045	0.00308	0.00067
Isced 2	0.03147	0.01395	0.04537	0.15087	0.08580	0.02904	0.21316	0.05109	0.16487	0.11158	0.09615	-0.01395	0.02067	-0.10754
	0.00018	0.00023	0.00161	0.00135	0.00024	0.00005	0.00283	0.00027	0.00032	0.00074	0.00147	0.00044	0.00310	0.00067
Isced 3	0.03848	0.01796	0.06911	0.15767	0.08605	0.06203	0.23137	Ref.	0.20227	0.11620	0.08637	0.03349	0.06037	-0.10528
	0.00017	0.00024	0.00162	0.00129	0.00025	0.00004	0.00285		0.00034	0.00071	0.00149	0.00046	0.00321	0.00069
Isced 5	0.09274	0.01220	0.06717	0.16348	0.08362	0.06404	0.24223	0.05372	0.18382	0.11591	0.11326	0.01092	0.05483	-0.08018
	0.00024	0.00033	0.00164	0.00137	0.00027	0.00006	0.00295	0.00028	0.00051	0.00075	0.00147	0.00047	0.00330	0.00076
Isced 6	0.02288	0.03110	0.09418	0.18442	0.09650	0.05772	0.26778	0.06657	0.20106	0.12525	0.13845	0.04916	0.07391	-0.10104
	0.00017	0.00025	0.00165	0.00135	0.00025	0.00005	0.00289	0.00029	0.00039	0.00073	0.00151	0.00045	0.00323	0.00076
Control variables														
Gender	-0.00788	0.02002	-0.00784	-0.00665	-0.00950	0.00538	-0.00903	-0.01022	-0.00927	-0.01460	-0.00291	-0.00601	-0.01211	-0.00401
(female = 1)	0.00004	0.00007	0.00009	0.00014	0.00002	0.00001	0.00012	0.00004	0.00004	0.00006	0.00013	0.00003	0.00019	0.00005
Age	-0.08135	-0.06446	-0.08767	-0.04136	-0.07450	0.01166	-0.02368	-0.08401	-0.09045	-0.06538	0.01451	-0.05624	-0.03284	-0.08418
	0.00012	0.00025	0.00026	0.00043	0.00006	0.00005	0.00034	0.00011	0.00013	0.00018	0.00039	0.00009	0.00052	0.00027
Migrant (=1	-0.07588	-0.08699	-0.09358	-0.03994	-0.06935	-0.13234	0.00540	-0.07246	0.01626	-0.04673	-0.05612	-0.04273	-0.03442	-0.10577
	0.00010	0.00024	0.00027	0.00035	0.00009	0.00004	0.00054	0.00016	0.00033	0.00022	0.00033	0.00013	0.00106	0.00042
Variances														
sigma_v	0.07536	0.04811	0.07114	0.05497	0.07245	0.06232	0.08232	0.05810	0.07903	0.06764	0.06333	0.05205	0.06791	0.05888
	0.00006	0.00008	0.00012	0.00015	0.00003	0.00002	0.00016	0.00004	0.00005	0.00007	0.00014	0.00003	0.00020	0.00009
sigma_u	0.32324	0.34986	0.24045	0.26971	0.19875	0.30821	0.27891	0.20901	0.40794	0.24712	0.27945	0.34697	0.36201	0.27835
	0.00009	0.00013	0.00019	0.00027	0.00004	0.00003	0.00025	0.00007	0.00008	0.00011	0.00024	0.00006	0.00036	0.00015
sigma2	0.11016	0.12472	0.06288	0.07577	0.04475	0.09888	0.08457	0.04706	0.17266	0.06564	0.08210	0.12310	0.13567	0.08094
	0.00005	0.00009	0.00008	0.00014	0.00001	0.00002	0.00013	0.00003	0.00006	0.00005	0.00013	0.00004	0.00025	0.00008
lambda	4.28949	7.27176	3.37974	4.90627	2.74335	4.94526	3.38791	3.59723	5.16207	3.65318	4.41264	6.66554	5.33044	4.72737
	0.00013	0.00017	0.00028	0.00036	0.00006	0.00004	0.00036	0.00010	0.00011	0.00017	0.00033	0.00008	0.00047	0.00021

Literacy seems to be strongly associated with the level of education. Across all countries, each level of education increases, on average, literacy by 37% in comparison to 'no schooling'. Not surprisingly, the increase in literacy increases with higher education levels. For example, a respondent who obtained a primary education degree (Isced 1) obtains, on average, 28% higher literacy than a respondent without schooling. Respondents with a second stage of tertiary education obtain, on average, 45% higher literacy scores than uneducated persons.

There are some interesting differences between more Anglo-Saxon countries – such as Canada, Ireland, USA, United Kingdom and New Zealand – and more continental European countries – such as Switzerland, Germany, the Netherlands and Belgium. In the Anglo-Saxon countries the average increase in literacy appears to be higher with increasing education level, than in continental European countries. Compare, for example, Canada and Germany. The average increase over education levels amounts to 38% in Canada, while only 14% in Germany. This difference arises from the (slightly) higher intercept term in continental European countries. This suggests a relative high average level of literacy in continental European countries with fairly small differences in literacy levels between education levels. In the more Anglo-Saxon countries average literacy levels are lower but differences between education levels are relatively larger suggesting that education contributes relatively more to one's literacy level than in the continental European countries. Surprisingly, although the average literacy level (observed from the intercept) is highest in the Netherlands, additional levels of education are estimated to be unfavourable to literacy. Street (1993), Gee (2008), and Cope and Kalantzis (2000) pointed to cross-cultural differences which might explain the difference in literacy. We leave the explanation of the cross-country literacy differences as scope for further research.

For parental education we find smaller but similar significant effect sizes: parental education – a proxy for socio-economic background – favours the literacy of the respondent. Pupils who have highly educated mothers are more educated themselves. This confirms previous findings in the literature (see Willms, 1999 for an overview).

Remarkable, in Switzerland (German speaking part) and Belgium the education of the mother has a higher impact on the literacy score than the respondent's education. In other words, the education of the mother determines the literacy of the respondent. This confirms previous findings that there is a large predeterminination in the Swiss and Belgian education system (Verschelde, 2009). A similar observation can be made for the education of the father in French speaking Switzerland. One can observe that in countries which experienced a large increase in wealth during the last decade (i.e., Ireland, Poland and Sweden), the education of the father was an important driver in literacy.

Gender differences in literacy levels appear to be small. However, in most countries we find a significant difference between male and female literacy levels. In all of these countries male literacy levels tend to be higher than female literacy levels. This confirms previous research by, among others, Charette & Meng (1997). Only in French speaking Canada and the USA, females have, controlled for own schooling and education of the parents, a higher literacy.

In most countries, except for the USA and New Zealand, we observe a negative relationship between age and literacy: i.e. literacy levels tend to be higher for younger people. In the USA and New Zealand, older persons have a better literacy than younger persons. This may indicate a shortfall in the education system. We did not find multiplicative effects with age [2].

Besides in Ireland and Poland, people who are born in their country of residence have a substantially higher literacy score than immigrants. Even in countries with a high immigration and an almost universal language – such as the USA – literacy among people not born in the country is substantially lower than literacy among people born in the country.

5.2 A Metafrontier Approach

The SFA model allows us to estimate the technical efficiency of observations (see Kumbhakar and Lovell, 2003). While accounting for the background characteristics in the estimation (which capture heterogeneity), the technical efficiency estimates reveal for each observation the shortfall in literacy in comparison to the best practice observation. For example, an efficiency score of 0.9 indicates that the respondent could increase his/her literacy score by 10%, given his/her background and schooling.

To estimate the shortfall in the production possibilities across countries, we follow the metafrontier framework of Battese & Rao (2002) and O'Donnell *et al.* (2008). As more formal discussion of the metafrontier can be found in the latter references, this paper is restricted to an intuitive outline. In a first step, a metafrontier framework pools all observations and estimates performance (denoted by μ^{meta}) against this pooled sample.[7] The potential best practice observation arises therefore not necessarily from the own group. The idea is that the most optimal transformation of inputs into literacy (the production technology) is, in theory, available for all countries. If country A has citizens who have, given their background, a higher literacy than in country B, country B shows a shortfall. The results of the metafrontier estimation are presented in Table 3. Sweden, Belgium and the Netherlands have, on average, the highest metafrontier efficiency. The metafrontier efficiency scores are, respectively, 0.879, 0.843 and 0.842, denoting that, if all students were performing as good as their (overall; i.e. from any country) best practice, average efficiency in generating literacy could increase in Sweden by 0.12% (Belgium 0.15% and Netherlands 0.15%). Given the background of the respondents, these three countries are performing better than the other countries in the sample. In sum, given the available resources, Sweden, Belgium and the Netherlands have an efficient school system in creating literacy. On the other hand, Poland and French speaking Canada have the lowest metafrontier efficiency score (scores of, respectively, 0.731 and 0.723). Their school system could significantly benefit from adapting it along the lines of the best practice countries.

Table 3 Efficiency estimates to group frontier and pooled metafrontier

	Relative to own frontier		**Relative to meta frontier**		
	Mean efficiency	St. dev. Efficiency	Mean efficiency	St. dev. Efficiency	Technology gap ratio
Sweden	0.8405	0.1010	0.8787	0.0972	1.0455
Belgium (Flanders)	0.8306	0.1095	0.8426	0.1051	1.0144
Netherlands	0.8625	0.0886	0.8423	0.0909	0.9766
New Zealand	0.8298	0.1103	0.8307	0.1104	1.0011
Germany	0.8605	0.0867	0.8292	0.0974	0.9637
Switzerland (French)	0.8365	0.1124	0.8213	0.1085	0.9818
Ireland	0.8166	0.1147	0.8138	0.1240	0.9966
Switzerland (German)	0.8442	0.0984	0.8046	0.1075	0.9531
Northern Ireland	0.7780	0.1446	0.8033	0.1435	1.0326
United Kingdom	0.7860	0.1354	0.7956	0.1300	1.0123
USA	0.7883	0.1426	0.7780	0.1437	0.9869
Canada (English)	0.7827	0.1290	0.7607	0.1462	0.9719
Poland	0.7541	0.1592	0.7313	0.1609	0.9698
Canada (French)	0.7630	0.1302	0.7233	0.1402	0.9480

[7] In fact, the pooled stochastic frontier analysis follows the line of Battese and Rao (2002). Later, Battese et al. (2004) solve a deterministic linear or quadratic programming problem to envelop the group frontiers. This paper follows the former approach.

In a second step, we first estimate the efficiency relative to all respondents in the same country. The resulting efficiency score is denoted by μ^{own}. De Witte and Marques (2010) interpreted the country average efficiency score as an indication for the inequality in learning possibilities in a country. The efficiency score is estimated as a relative efficiency score (i.e., relative to all other observations in the reference set; *in casu*, the respondents of a country). In the SFA model, the reference set is restricted to observations with similar characteristics, as proxied by the control variables. If a respondent with similar characteristics is obtaining higher efficiency scores, the SFA model will split this difference to a stochastic error term (capturing the random noise) and an inefficiency term (capturing how much the respondent could do better in terms of literacy score, if he/she would perform as efficient as his/her comparable best practice). The higher the average efficiency score of a country, the more equal the opportunities for learning, given the control variables. The results are presented in Table 3. The average efficiency amounts to 0.86 in the Netherlands and Germany. These countries are best able to provide for similar respondents similar literacy rates (as was also observed from the eye-ball econometrics in the previous section). On the contrary, Poland (score of 0.754), French speaking Canada (0.763) and Northern Ireland (0.778) do not succeed in giving their citizens similar opportunities, given their background (cf. previous section).

Finally, in a third step, we compare the outcomes of the previous two steps. By dividing the meta efficiency score μ^{meta} by the regular score μ^{own}, we obtain the technology gap ratio (TGR). TGR estimates the shortfall in efficiency by not adapting the best available 'technology' for transforming the inputs into literacy (see Battese & Rao, 2002; O'Donnell *et al.*, 2008; and De Witte & Marques, 2009). The TGR proxies the potential efficiency gains by changing the educational system and learning from the best practice countries (i.e., the countries with a higher meta efficiency score μ^{meta}). Due to sample size bias (i.e., the larger the reference sample, the lower the average efficiency; Zhang and Bartels, 1998), μ^{meta} is generally lower than μ^{own}. Consequently, the TGR will be smaller than 1 (although this is not necessarily the case as the SFA approach controls for random noise), pointing to potential efficiency gains. The results in Table 3 reveal that particularly French speaking Canada (TGR = 0.948), Switzerland (German part; TGR = 0.953) and Poland (TGR = 0.970) could benefit from learning from the best practice countries.

6. Conclusion and Discussion

The growing importance attached to education coupled with the continuous scarcity of resources for education make it important that skills and literacy are produced efficiently. This paper provides an international comparison of the efficiency of literacy and finds substantial differences between countries in levels of literacy, differences in literacy between education levels and differences in the efficiency of literacy production. There are some notable differences between more Anglo-Saxon countries and the Continental European countries. Typically, Anglo-Saxon countries – like the USA, Great Britain, Canada and Ireland – have lower levels of literacy but larger differences in literacy levels between levels of education than continental European countries such as Sweden, Germany, Netherlands and Belgium.

On the other hand, the estimates suggest that in continental European countries as Sweden, Belgium and the Netherlands literacy is produced more efficiently than in countries like the USA, Canada and Poland. So, it seems that large differences in literacy levels between levels of education are associated with a relatively higher efficiency of literacy production.

Finally, the findings suggest that in almost all countries the scope for efficiency improvements in education is large (i.e., we observe average performance scores which are significantly smaller than 1.0). It appears that even without major increases in (public) funding for education, improvements in educational outcomes may be achievable by learning from the best practices.

Further qualitative research should elaborate on why some countries (as Netherlands or Sweden) are able to generate with the given resources and background of the students more literacy than other countries (as Poland and Canada).

Acknowledgments: We are grateful to seminar participants at University of Groningen for insightful comments and discussions. The usual caveat applies.

References

[1] Barbetta, G. and G. Turati (2003), "Efficiency of junior high schools and the role of proprietary structure", *Annals of Public and Cooperative Economics*, 74(4): 529-552.

[2] Battese, G. and D. Rao (2002), "Technology gap, efficiency and a stochastic metafrontier function", *International Journal of Business and Economics*, 1(2): 87-93.

[3] Battese, G., D. Rao and C. O'Donnell (2004), "A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies", *Journal of Productivity Analysis*, 21(1): 91-103.

[4] Becker, G.S. (1992), "Human Capital and the Economy", *Proceedings of the American Philosophical Society*, 136(1): 85-92.

[5] Blau, F. & L. Kahn (2005), "Do cognitive test scores explain higher US wage inequality?" *Review of Economics and Statistics*, 87(1): 184-193.

[6] Blum, A., H. Goldstein and F. Guérin-Pace (2001), "An analysis of international comparisons of adult literacy", *Assessment of Education: Principles, Policy and Practice*, 8(2): 225-246.

[7] Charette, M.F. and R. Meng (1997), "The determinants of literacy and numeracy, and the effect of literacy and numeracy on labour market outcomes", *Canadian Journal of Economics*, 31(3): 495-517.

[8] Cope, B. and Kalantzis, M. (2000), *Multi-Literacies: Literacy Learning and the Design of Social Futures*, K. Psychology Press.

[9] Denny, K., C. Harmon, and S. Redmond (2000), "Functional literacy, educational attainment and earnings – evidence from the international adult literacy survey", *IFS Working Paper* W00/09.

[10] De Witte, K. and R. Marques (2009), "Capturing the environment, a metafrontier approach to the drinking water sector", *International Transactions of Operational Research*, 16(2): 257-271.

[11] De Witte, K. and R. Marques (2010), "Designing performance incentives, an international benchmark study in the water sector", *Central European Journal of Operations Research*, 18(2): 189-220.

[12] Hanushek, E. (1979), "Conceptual and empirical issues in the estimation of educational production functions", *Journal of Human Resources*, 14(3): 351-388.

[13] Hanushek, E. (1986), "The economics of schooling: Production and efficiency in public schools", *Journal of Economic Literature*, 24(3): 1141-1177.

[14] Hanushek, E. and J. Luque (2003), "Efficiency and equity in schools around the world", *Economics of Education Review*, 22(5): 481-502.

[15] Gee, J.P. (2008), *Social Linguistics and Literacies: Ideology in Discourses* (3rd Ed.), London and New York: Taylor and Francis Group.

[16] Izadi, H., G. Johnes, R. Oskrochi and R. Crouchley (2002), "Stochastic frontier estimation of a CES cost function: the case of higher education in Britain", *Economics of Education Review*, 21(1): 63-71.

[17] Jondrow, J. C. Lovell, I. Materov and P. Schmidt (1982), "On the estimation of technical inefficiency in the stochastic frontier production function model", *Journal of Econometrics*, 19(2,3): 233-238.

[18] Kumbhakar, S.C. and C.A.K Lovell (2003), *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press, p. 333.

[19] Newhouse, J. (1994), "Frontier estimation: how useful a tool for health economics?" *Journal of Health Economics*, 13(3): 317-322.

[20] Meeusen, W. and J. Van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", *International Economic Review*, 18(2): 435-444.

[21] O'Donnell, C., D. Rao and G. Battese (2008), "Metafrontier frameworks for the study of firm-level efficiencies and technology ratios", *Empirical Economics*, 34(2): 231-255.

[22] OECD (2000), "Literacy in the information age – Final report of the international adult literacy survey", *Directorate for Education*, [On-line] Available: http://www.oecd.org/LongAbstract/0,3425,en_2649_39263294_39437981_1_1_1_1,00&&en-USS_01DBC.html

[23] OECD (2006), "Education at a Glance", *Directorate for Education*, [On-line] Available: http://www.oecd.org/document/52/0,3746,en_2649_39263238_37328564_1_1_1_1,00.html

[24] Verschelde, M. (2009), "Estimating and explaining efficiency of secondary education", *University of Gent, Working paper* series, No.17: 1-49.

[25] Willms, J.D. (1999a), "Literacy skills and social class", *Options Politique*, 7(8): 22-27.

[26] Willms, J.D. (1999b), "Quality and inequality in children's literacy – The effects of families, schools and communities". In Keating, D.P. & C. Herzman (Eds.), *Developmental health and the wealth of nations: social, biological and educational dynamics*. New York: The Guilford Press: pp. 406.

[27] Ruggiero, J. and D. Vitaliano (1999), "Assessing the efficiency of public schools using data envelopment analysis and frontier regression", *Contemporary Economic Policy*, 17(3): 321-331.

[28] Schultz, T. (1967), *The Economic Value Education*, New York: Columbia University Press.

[29] Street, B.V. (1993), Cross-cultural approaches to literacy, Cambridge: Cambridge University Press.

[30] Zhang, Y. and R. Bartels (1998), "The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand", *Journal of Productivity Analysis*, 9(3): 187-204.