Efficiency in Public Education

Waldo, Staffan

2002

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Efficiency in Public Education

Staffan Waldo*
Department of Economics
Lund University
P.O. Box 7082
SE-220 07 Lund
Sweden
staffan.waldo@nek.lu.se
Keywords: Data Envelopment Analysis, Education
JEL: I21

February 28, 2002

Abstract

The purpose of the paper is to investigate the influence on efficiency in public education from teacher characteristics and private school competition. Using Data Envelopment Analysis (DEA) an efficiency model introduced by Olesen & Petersen (1995) is extended to include the production environment of the schools. This increases the number of dimensions in the DEA why the possibility to discriminate between production units using educational data is discussed. Efficiency is estimated for 851 of Sweden's approximately 1000 secondary schools in 1994/95. Estimated efficiency implies that the schools can use 9-19% less resources on average, still achieving the same results. The efficiency scores are used as dependent variable in a Tobit regression model. Our primary finding is that competition from private schools has a positive influence on the efficiency in public education but that the teacher characteristics do not significantly explain any efficiency differences.

1 Introduction

The influence of school resources on educational outcomes has been studied rigorously for the last decades. By regression analysis the relation between 'knowledge' and variables

*I would like to thank Lars-Gunnar Svensson, Magnus Wikström and Pontus Roos for useful comments, to Curt Wells for providing Gauss routines for the statistical models and to participants at seminars where the paper has been presented. Data is provided by the Swedish National Board for Education. The project is financed by the Jan Wallander foundation.

from school, family and society is explained. The results are often weak and point in different directions. The most stable result is that family background is the most important factor in explaining educational results. Children with for example highly educated parents will in general do better in school.

One of the basic question is how to provide an optimal amount of education using a minimum of resources. Despite the weak results from economic research the importance of teachers' experience, teachers' education etc is put forward in the public debate. On the other hand the influence from private schools on the public school system is more criticized. In economics the advantages of competition to the public provision is often emphasized in both theory and empirical analysis. This paper analyses empirically the role of teachers and private school competition for efficiency in the provision of public education.

When estimating efficiency in public education the use of Data Envelopment Analysis (DEA) has a number of advantages, e.g. no prices are needed and it is possible to model both input and output in multiple dimensions. The flexibility of the method makes it possible to model a production process incorporating the specific characteristics associated with education. This study departs from a model originally developed by Olesen & Petersen (1995) using relative grades as output. The students attending a school will during the process achieve certain skills (operationalized as grades) using the inputs offered by the school. Since the grades are relative and thus contain no information on the absolute difference in knowledge their ability to discriminate between different distributions of the grades is limited which is taken into account in the model. Output is modelled as cumulative frequencies of students reaching the different quality levels. We extend the model by incorporating the socioeconomic status of the students as differences in the production environment. By including the number of students attending the schools as an input in the models we indirectly include drop out rates as a performance measure since only students with registered grades are modelled as output. The discriminatory power of the model is discussed in relation to e.g. Muñiz (n.d.). The estimated efficiency is used as dependent variable in a Tobit regression model when determining the effects of competition and teacher characteristics on efficiency.

A number of studies have used efficiency models to measure school performance but with other objectives than explaining why some schools are more efficient than others.[1]

[1] Ljunggren (1999), Grosskopf, Hayes, Taylor & Weber (1997), Färe, Grosskopf, Försund, Hayes & Heshmati (1997), Olesen & Petersen (1995), Grosskopf, Hayes, Taylor & Weber (1999), McCarty & Yaisawarng (1993), Färe, Grosskopf & Weber (1989) and Heshmati & Kumbhakar (1997).

Studies trying to explain efficiency are e.g. Kirjavainen & Loikkanen (1998), Duncombe, Miner & Ruggiero (1997), Lovell, Walters & Wood (1994), Bradley, Johnes & Millington (2001), Grosskopf, Hayes, Taylor & Weber (2001) and Waldo (2000). The studies are from different countries and evaluate different levels of education, but they indicate that there are factors explaining efficiency differences both in the market structure, the political context and in the internal organization of the schools.

This study estimates efficiency in Swedish secondary schools and use the estimated efficiency scores as dependent variables in a regression analysis. The data consists of 851 Swedish secondary schools in 1994/95. Education in Sweden is primarily financed by municipal income taxes and governmental grants. In 1991/92 the grant system for education in Sweden was changed from a direct subsidy to a lump sum system where education competes for funds with other municipal activities. One of the ideas behind the change was to make governmental activities more efficient, and the possibilities to organize schools to fit local conditions were increased. At the same time competition from private schools was introduced to the educational system.

The most discussed school input is the teachers, not surprising since teachers are the main resource at a school. They work close to the students, which gives them great influence both on the learning process and the results. The influence of teachers on the results is far from determined. E.g. Hanushek (1996) concludes in a survey of the literature on education that there is no clear positive relationship between teaching resources and student outcome. However, using a sample of studies Hedges & Greenwald (1996) come to the opposite conclusion using statistical methods. For Swedish conditions the National Board for Education has studied the relationship using a data set containing all Swedish secondary schools (Skolverket (1999)) where the influence of teachers on mean grades is not clear. In our study non of the teacher characteristics continuity, experience and pedagogical skill are significant over a set of Tobit models. We thus find no evidence of differences among the teachers in public schools having any impact on the possibilities to provide education with a low resource use.

From an economic view competition is highly interesting and e.g. Grosskopf et al. (2001) find competition to improve allocative efficiency but not technical efficiency and Bradley et al. (2001) find a positive relationship between technical efficiency in English secondary schools and competition from non selective schools in their neighborhood. On the other hand Duncombe et al. (1997) find that competition decreases cost efficiency in New York school districts and Kirjavainen & Loikkanen (1998) find that private schools are less efficient than public in Finland. A positive influence on public school efficiency from private school competition is the single most significant result in our analysis.

The paper continues with section 2 where the basic concept of Data Envelopment Analysis is outlined. In section 3 the data and models used in this paper are specified and the results from the efficiency estimation are presented. Section 4 contains the analysis on what variables explain the estimated efficiency and in section 5 the results are discussed. Section 6 is a summary.

2 Estimating efficiency

2.1 Data Envelopment Analysis

When evaluating the performance of a production unit we are interested in the relation between inputs and outputs. A unit performing well uses a small amount of inputs in order to produce a large amount of outputs. When estimating the performance with DEA we look at units with the same amount of outputs and compare their input quantities. This is referred to as an input oriented model.[2]

In the DEA literature the performance measure is called *technical efficiency*. Input oriented technical efficiency measures if a production unit actually uses the minimum possible amount of inputs necessary to produce the observed output. The idea is best shown in a simple figure, see figure 1.

Figure 1: About here

The figure shows four production units, a,b,c and d, producing one output using one input. What units are efficient depends on what we believe about the production process. One possibility is to assume that we have a constant returns to scale (CRS) technology. In this case we can scale any possible production set (x, y) up or down proportionally. Unit b is efficient in the CRS technology and the production observed for unit b can be scaled up or down along the CRS frontier. The other production units are compared to the CRS frontier. If they produce 10% less output they should be able to use 10% less input. For example unit a is inefficient compared to the CRS frontier since it should be able to produce at a', a point on the CRS frontier having the same output but less input. If we do not think it is possible to scale production this way we could use a variable returns to scale (VRS) technology. This technology envelops the data closer than does

[2] For a thorough analysis of efficiency measures see Färe, Grosskopf & Lovell (1994) .

the CRS technology and we define units a,b and c as efficient in the diagram. Unit d is still inefficient. According to the CRS frontier unit d should be able to produce at d' and according to the VRS technology unit d should be able to produce at d". d' is a radial contraction of the input/output combination observed at b, whereas d" is a linear combination of observations a and b.

Figure 1 illustrates a production with a single input and a single output. One of the advantages with the DEA method is that it can handle a multiple input and multiple output production process. With two or more inputs in an input oriented model, efficiency is measured as the possibility to decrease the inputs radially, i.e. decreasing all with the same proportion.

What we try to find is the smallest possible constant, λ, with which we can multiply all inputs and still be within the observed production possibilities. λ is our measure of efficiency and we will get an efficiency score between zero and one. The smaller λ the more it is possible to decrease input and the more inefficient is the investigated unit.

We define the production possibilities as the set of possible input vectors, x, that can produce the output vectors u as

$$L(u|C,S) = \{x : u \leqq zM, x \geqq zN, z \in \Re^K_+\} \tag{1}$$

where M is a matrix of outputs, N is a matrix of inputs, z is a vector of activity variables and K is the number of observations. C refers to a CRS technology and S to strong disposability of inputs[3]. Holding the output vector constant $L(u|C,S)$ can be shown graphically as in fig 2.

Figure 2: About here

The units a and b are on the efficient frontier but c is inefficient since the unit can contract the inputs and still be inside $L(u|C,S)$. Contracting unit c's inputs with λ will generate a production on the frontier. λ is the smallest possible constant with which we can multiply unit c's input and still be inside $L(u|C,S)$.

The measure of technical efficiency, λ, is then

$$F_i(x^k, u^k|C,S) = \min\{\lambda : \lambda x^k \in L(u^k|C,S)\}, k = 1, 2, ..., K \tag{2}$$

Efficiency for unit k is estimated with the following linear programming model.

[3]The strong disposability assumption implies that an increase of an input does not decrease output.

$$
\begin{aligned}
F_i(x^k, u^k|C,S) &= \min \lambda \qquad (3)\\
s.t. \sum_{k=1}^{K} z^k u_m^k &\geqq u_m^k, m = 1, ..., M\\
\sum_{k=1}^{K} z^k x_n^k &\leqq \lambda x_n^k, n = 1, ..., N\\
z^k &\geqq 0, k = 1, ..., K
\end{aligned}
$$

Imposing the restriction that the z variables sum to one we get the VRS model, without the restriction we have a CRS model.

As mentioned above the most important variables for school performance is the students' socioeconomic background. These variables are not under the control of the schools, at least not in the short run, and one way of measuring efficiency is to include them in the model creating a measure of efficiency that is not due to differences in socioeconomic background. The socioeconomic variables have to be treated separately since the efficiency measure is a contraction of the input variables and we cannot contract the socioeconomic inputs. The solution is to estimate a model where there is no possibility of contraction, i.e. no λ for the vector of socioeconomic variables. Instead these variables represent the production environment and we only compare units having students with similar socioeconomic status. We define the technology as

$$L(u|C,S) = \{(x_\alpha, x_{\hat{\alpha}}) : u \leqq zM, x_\alpha \geqq zN_\alpha, x_{\hat{\alpha}} \geqq zN_{\hat{\alpha}}, z \in \Re_+^K\} \qquad (4)$$

where x_α represents the N_α inputs that can be contracted and $x_{\hat{\alpha}}$ represents the $N_{\hat{\alpha}}$ inputs that are quasi fixed. The efficiency measure thus does not contract the $x_{\hat{\alpha}}$ inputs

$$F_i(x_\alpha^k, x_{\hat{\alpha}}^k, u^k|C,S) = \min\{\lambda : (\lambda x_\alpha^k, x_{\hat{\alpha}}^k) \in L(u^k|C,S)\}, k = 1, 2, ..., K \qquad (5)$$

The DEA framework outlined above will be used for estimating the performance of Swedish secondary schools. The inputs, outputs and socioeconomic variables used in the models are discussed in section 3.

3 Data and Models

At the time of the school reform in 1991/92 the Swedish National Board for Education started collecting data from all Swedish municipalities about school resources, socioeconomic background facts on the students and educational results. The data is used by the national government and the municipalities to compare education in the municipalities. The data set has been disaggregated into school level and the data set contains a number of student and school characteristics for all Swedish secondary schools. The Swedish National Board for Education (Skolverket (1999)) use it in a regression model with mean grades as output. It is focused on the teaching resource, but the influence in the analysis is not clear. In their analysis they find that parents' education is the single most important factor explaining grades. Other socioeconomic variables of importance are immigrant students and sex.

The same data is used for the present study with a few adjustments. Schools with less than twenty students graduating are discarded. We are interested in getting results valid for some kind of 'average' school and schools with less than twenty students are often either located in very sparsely populated areas or specialized towards some special group of students. We have also discarded schools with bad or missing data. Altogether 851 schools are included in the study (in 1994/95 Sweden had 1062 secondary schools).

3.1 Input Data

The teaching resource is the main resource used to increase students' knowledge. As the first input in the models we use *teaching hours*, which is the number of ordinary teaching hours provided to the students each week at a school. Schools can also provide *special needs teaching* for students needing extra teaching hours, which is our second input. These two inputs are thought to be under direct administrative control of the individual school, which is our investigation unit, and are therefore treated as discretionary inputs in the models.

Number of students is included as a nondiscretionary resource in some of the models. The students are here viewed as a necessary 'resource' at a school such that without students there will be no grades as output. The number of students is non discretionary when having the individual school as investigation unit since the school cannot choose how many students to educate. This is rather determined by the age structure of the population together with political decisions. Students are of course not a homogeneous group but differ in their ability to achieve knowledge. Well known from the literature on education is that the socioeconomic status (SES) of the student is correlated with educational results. We define students socioeconomic status by if they are immigrants and by their parents' education. Parents' education is the most important socioeconomic resource as discussed above. The better educated parents a student has, the better she performs on average. A student with a highly educated parent is defined as having at least one of the parents with a university degree. Other students have parents with low education. Immigrant students do on average achieve lower grades. Immigrant students are defined as students either born outside Sweden or with both parents born outside Sweden.

When controlling for differences in the student-body we divide the students into three categories: *SES group A* contains Swedish students with highly educated parents, *SES group B* contains Swedish students with low educated parents and immigrant students with high educated parents while *SES group C* contains immigrant students with low educated parents.

3.2 Output Data

As output from the educational process we have grades for *Swedish language, mathematics and English.* These are regarded as the three most important subjects in Swedish education and the grades are also based on national tests taken by all students which makes them better to compare. In each subject it is possible to achieve grade 1,2,3,4 or 5, with 5 being the highest.

Our output measure follows a model for school efficiency developed by Olesen & Petersen (1995). Knowledge is operationalized as the number of students achieving different quality levels, i.e. different grades. A school will thus produce y_1 number of students achieving grade one, y_2 students achieving grade two etc. However, viewing each quality level as a separate output dimension makes it possible to be efficient by specializing at one quality level. A school producing only students at a specific quality level could be defined as efficient because it is extreme in that dimension, even though it performs poorly at all other output qualities. To control for this problem we follow Olesen & Petersen (1995) and use cumulative frequencies where the cumulative frequency for knowledge level one is all students with grade one or better, the cumulative frequency for knowledge level two is all students with grade two or better etc.

Since the output measure is not standard in the literature we give a simple example to explain the basic concepts[4]. Consider three different schools having students graduating with grades from one to five as illustrated in table 1.

The number of students who have achieved the different quality levels at each school

[4]For a further discussion of the model see Olesen & Petersen (1995).

School		1	2	3	4	5	Total
A	Freq	2	2	2	2	2	10
	Cum freq	10	8	6	4	2	
B	Freq	1	2	4	2	1	10
	Cum freq	10	9	7	3	1	
C	Freq	5	0	0	0	0	5
	Cum freq	5	0	0	0	0	

Table 1: Output example

is presented in the row 'Freq' and the corresponding cumulative frequency is presented in 'Cum freq'. Comparing school A and B show that school B has higher cumulative frequency in grades 2 and 3, but school A has higher in grades 4 and 5. School A thus has more students with both low and high achievement which is clear from the row showing frequency. We cannot define that either A or B has more output than the other since we cannot state that the low performing students at school A are compensated by the high performing students. School C is rather poor performing since all students have the lowest grade. Using cumulative frequencies both A and B have larger output at all quality levels. However, if measuring output as frequencies there would be no school dominating school C because of the large output in grade 1.

3.3 Model Specification

DEA model I follow the idea of Olesen & Petersen (1995) in measuring the transformation of school resources directly into a cumulative outcome of students graduating. This measures for the individual school the teaching resource necessary for reaching the observed number of students and their distribution among the possible grades. But it does not take into account the number of students actually attending the school. When evaluating an inefficient unit the point on the front might have a larger number of students attending school, but loose some of them in the process. Thus the dropout rate is not taken into account in the model.[5]

Model II and III are built around the students as an input. The idea is that schools produce knowledge and without students as inputs it will not be possible to produce any output. Since the number of students is determined outside the individual school, they are

[5] Olesen & Petersen (1995) restrict their analysis to students passing an exam and continue by defining the *probability* of achieving different levels of knowledge or better (i.e. cumulative grades) as output.

treated as a non discretionary input. By, as in model II, including the number of students in the model, the school under evaluation will have at least as many students attending school as the frontier. Since the frontier must have at least as much output (students) in all dimensions, including the whole range of grades, they are also forced by the model to have at least as many students as output. Thus in model II an inefficient school will have at least as many students not passing as the frontier. In model III the socioeconomic status of the students is taken into account. A unit defined as inefficient in the previous models might be so because a large share of the students have low socioeconomic status and are thus more resource demanding. In model III students from SES group A and B are treated as above, while the most resource demanding students in SES group C are forced to be equal for the frontier and the school under evaluation. In this way the inefficient schools will not have a larger share of the students belonging to SES group C. The models are presented in table 2 where X represent an input, Y an output and XY is a variable that is not defined as either, i.e. it is forced to be equal for the school under evaluation and the frontier.

School resources	**Model I**	**Model II**	**Model III**
Ordinary teaching	X	X	X
Special needs teaching	X	X	X
Nondiscretionary resources			
Students - total	-	X	-
SES group A	-	-	X
SES group B	-	-	X
SES group C	-	-	XY
Results			
Cumulative grades 1-5 in Swedish	Y	Y	Y
Cumulative grades 1-5 in English	Y	Y	Y
Cumulative grades 1-5 in Mathematics	Y	Y	Y

Table 2: DEA Model Specification

As discretionary school inputs we use teaching hours in ordinary teaching and special needs teaching.[6] The number of students attending a school is included in model II and divided into the socioeconomic categories discussed above in model III. Output is measured as cumulative frequencies at different achievement levels for Mathematics, Swedish

[6]Note that input is the teaching resource for all subjects while output only contains three of them.

and English.

The models are estimated using input orientation for both a CRS and a VRS frontier. In the efficiency estimation a large number of units are defined as efficient as presented in Table 3. With many input and output dimensions a unit may have a number of inputs and/or outputs that differ from other units and we might not find another unit to compare with. Such a unit will be defined as efficient (but with no other references than itself), not necessarily because it is better than others, but because it is different. Table 3 shows the total number of efficient schools and the number of them with and without references.

	Total	**Without ref**	**With ref**
MI CRS	33	4	29
MI VRS	54	10	44
MII CRS	78	27	51
MII VRS	123	54	69
MIII CRS	289	126	163
MIII VRS	389	203	186

Table 3: Efficient Schools

In the analysis below the efficient schools without references are discarded from the data. We do not want mean efficiency to increase when including socioeconomic variables due to an increase in the number of units differing from the others.[7]

Mean efficiency estimated in the models is shown in table 4. Efficient units get a DEA score of one while inefficient units get a score less than one. A DEA score of e.g. 0.8 implies that the unit should be able to reduce inputs with 20% without having to decrease outputs.

For the model not including students mean inefficiency is approximately 17-19% depending on the returns to scale assumption. Including the students the mean inefficiency is 16-17% while including socioeconomic variables decreases mean inefficiency to 9-10%.

[7]The number of efficient units increase rapidly with the inclusion of SES variables. The low discriminatory power of models including environmental variables is discussed in e.g. Muñiz (n.d.). Muniz suggests an alternative way of modelling SES based in a three stage model where the number of efficient units do not increase in the same way and thus the discriminatory power remains high. The models proposed in this paper have a somewhat different approach beeing restrictive with discriminating between schools. This is because of the nature of the educational data used which only approximates what is really to be measured.

	CRS	VRS
MI	0.812	0.826
MII	0.829	0.842
MIII	0.902	0.913

Table 4: Mean Efficiency

The inclusion of socioeconomic variables in the model increases efficiency radically indicating that a large proportion of the estimated inefficiency in the other models are due to differences in the production environment of the schools. This is consistent with the literature on student performance where the role of socioeconomic background of the student is an important result. In the continued analysis below we concentrate on efficiency estimated in the DEA models either including the socioeconomic status of the students or not taking students into account at all, i.e. models I and III. We use the model estimated under variable returns to scale since this is the most flexible assumption[8]

4 Explaining Efficiency

The focus of this study is on school performance and if efficient performance can be explained by the internal organization or structures of the school environment. Below variables concerning internal characteristics and production environment of the schools used for explaining efficiency differences are discussed, with special attention to competition and teacher characteristics. These are used as independent variables in a regression analysis. We thus follow a two step procedure by first modelling the production of education and estimating efficiency and then modelling efficiency as dependent on a number of factors not included directly in the production model.

The second stage analysis differ somewhat between the efficiency models. Model III estimates efficiency taking environmental differences into account. Increasing the number of dimensions increases the number of efficient units and, as seen in table 3, also the number of isolated efficient units. In the second stage model III has less observations than the other models and a larger number of limit observations (i.e. fully efficient). Including the environmental variables in the efficiency model will make them unavailable for the second stage regression. In this way there is a trade off between the models in the two stages since the regression model will 'loose' highly significant environmental

[8]Using constant returns to scale does not change the qualitative analysis below.

variables, the number of observations decreases and the dependent variable show less variation. An unnecessarily strict definition of the environment might thus imply a loss of important information in the sample. In model I efficiency is estimated with no concerns taken to environmental variables. These are used as explanatory variables in the second stage. By reducing the number of dimensions the discriminatory power of the model will increase. In the efficiency estimation a school with low student SES and thus more resource demanding students will in many cases be compared with schools having high student SES. The difference in efficiency due to the production environment is explained in the second stage regression model where the sample in this case will contain more information than for model III due to more observations of which less are defined as efficient.

Turning to the explanatory variables, competition from *private schools* may influence the public schools in different directions. In economic theory competition is often seen as forcing firms to be efficient in order not to be put out of the market. This would imply that public schools facing competition will be more efficient than other schools. But private schools might also compete for the most motivated students, thus leaving the public schools with students less interested in school work. This scenario will of course have a negative impact on the public schools. Which effect is dominating is an empirical question. E.g. Bergstrom & Sandstrom (2001) have found a positive relationship between competition from private schools and student outcomes in public schools in an empirical study on Swedish students graduating from secondary school while Waldo (2000) find no significant relation between the proportion of students in private schools and efficiency in the provision of education by Swedish municipalities. Defining private school competition as the share of the municipal students that attend a private school is an attempt to measure the degree of competition. The market share of dominating firms is a common approximation of competition in empirical studies. However, a large share of private school students might also reflect a bad response to competition from the public schools.

Teachers are important since they work close to the students and constitute the main resource that the school can offer the students in their learning process. The teaching resource is used as input in the DEA models. But teachers is not a homogeneous input and teachers have different abilities to increase the knowledge of the students. We have a number of characteristics that might influence the teaching ability of the teacher. The first is *experience*. A more experienced teacher should have achieved a certain skill in teaching his or her subject. On the other hand a teacher coming directly from her own education might be more enthusiastic and eager to teach and also better skilled in modern teaching methods. The actual influence will be an empirical question. *Teaching experience* is operationalized as the mean number of years the teachers at a school have been registered as teachers. Teaching *continuity* is thought of as positively influencing the students' results since a class facing many different teachers in a subject must also become acquainted with the teaching methods of the new teacher and the teacher must come to know the students to be able to help them in the best possible way. *Teaching continuity* is here defined as the proportion of teachers that remain with the student throughout secondary school. *Pedagogical skill* is regarded as an important tool for improving educational outcomes. Much effort is spent on increasing the pedagogical skills of the teachers, but schools often have problems finding pedagogically educated teachers. When estimating efficiency we expect better educated teachers to increase efficiency since if their pedagogical skills makes any difference in the learning process, their students should be able to learn more than the students of unskilled teachers putting the same effort into teaching. *Pedagogical skill* is defined as the part of the teachers at a school having formal pedagogical skills in teaching.

The location of the schools is examined by a dummy variable, *Large city*, with one indicating schools situated in urban areas and zero indicating all other schools.

Popular in local school politics in Sweden is merging primary and secondary schools with the idea that mixed schools contribute to the development of the school children. We test the idea that *pure secondary* schools are less efficient in providing education. Pure secondary schools are included in the analysis as a dummy variable having one as a pure secondary school and zero as all other schools.

The data is collected by the Swedish National Board for Education in the same data set as discussed above[9]. Descriptive statistics of the explanatory variables are shown in table 5.

4.1 Models and Results

The Tobit model is the most common model in the literature when explaining efficiency and is used by i.e. McCarty & Yaisawarng (1993) and Kirjavainen & Loikkanen (1998) for school data. The Tobit model is specified as follows

[9]Except for Large City which is defined for all municipalities by the Swedish Bureau of Statistics (SCB) and competition from private schools which is originates from the data base "Jamfortal for skolhuvudman", administred by the Swedish National Board for Education.

	Mean	St dev	Min	Max
Private schools	1.826	2.381	0	9.6
Teachers' experience	16.883	2.749	2.3	24.1
Teachers' continuity	72.336	10.517	23.53	100
Teachers' skill	0.923	0.061	0.528	1
Large City	0.350	0.477	0	1
Pure secondary	0.428	0.495	0	1

Table 5: Explanatory Variables

$$
\begin{aligned}
eff_i &= \beta' X_i + u_i \quad if\ eff < 1 \\
&= 1 \qquad\qquad if\ eff \geq 1
\end{aligned}
$$

Where eff_i is efficiency score for municipality i, X_i is the explanatory variables and u_i is a normally distributed residual with mean 0 and variance σ^2.

In Table 6 a Tobit model for the efficiency scores in models I and III are estimated.[10]

Presented in Table 6 are coefficients and not the marginal effects since we are only concerned with the sign of the variables. The only variable that is significant in both models is competition from private schools where a large share of the students attending a private school implies more efficient public schools. This would be consistent with a hypothesis of private schools (maybe consisting of schools with different educational profiles) entering the market and that public schools loosing students to private schools have been forced to increase efficiency. Teacher continuity and experience are insignificant in both models. Schools with high teacher skill and pure secondary schools are more efficient when not taking the production environment into account in the efficiency estimation in model I. The relationship vanishes in model III. Schools in large cities are more efficient in model III but has no significant effect in model I. The environmental variables in the

[10]Normality is tested with a conditional moment test, see e.g. Greene (1997), where for Model I LM=14.48 implies a non normal distribution (The test distribution is Chi2 with 2 df) and for Model III LM = 4.67 implies normality. Heteroscedasticity is found in a LR-test for Model I, see e.g. Greene (1997), where LR=14.69 (Chi2 with 8 df) and for Model III hwere LR = 16.88 implying that the models are heteroscedastic. The Tobit models are therefore estimated with heteroscedasticity of the form

$$\sigma_i = \sigma \exp(\gamma z_i)$$

where z consists of the same variables as X.

Dep variable:	Log of efficiency scores			
	MI		**MIII**	
	Coeff	**p-value**	**Coeff**	**p-value**
Const	-0.6066	0.0000	-0.0905	0.2292
Private schools	0.0110	0.0000	0.0088	0.0001
Continuity	0.0004	0.1737	0.0007	0.1509
Experience	-0.0005	0.7418	-0.0009	0.6686
Skill	0.1525	0.0227	-0.0524	0.5659
Large city	0.0097	0.2344	0.0290	0.0176
Pure secondary	0.0182	0.0088	0.0010	0.9144
Parents' education	0.1277	0.0000	-	-
Immigrants	-0.0033	0.0000	-	-
Specification test				
Normality	LM = 14.42	0.0007	LM = 4.67	0.0969
No of units	841		648	

Table 6: Tobit Model Explaining the Efficiency Distribution in Model I and Model III (VRS)

model I regression are both significant and have the expected signs.

5 Discussion

With limited resources it is necessary to provide an efficient public education in order to reach educational goals. The best way of doing this is widely discussed. This study focuses on two subjects common in the debate, competition from private schools and teacher characteristics.

In the statistical analysis we find a positive influence on efficiency from private school competition. The finding is consistent with Hoxby (2001) and with Bergstrom & Sandstrom (2001) where Swedish students achieved higher grades and test scores in schools with more private competition, but not with Waldo (2000) where private school competition had no significant influence on efficiency in the provision of education by Swedish municipalities. The conclusion from this study is that public schools facing competition do use their teaching resources more efficient in the provision of education to the citizens.

In the literature the influence on school results by teacher characteristics is far from determined. Many studies get insignificant results which can be interpreted as teacher characteristics not improving school outcomes. However, as Hedges & Greenwald (1996) point out, many of the studies show a positive relationship which they interpret as that although the individual studies are not significant the studies together show a positive significant relationship. If teacher characteristics have a positive influence on school results, we expect schools with 'better' teachers to be more efficient since they can reach the same result by using less resources. We find no evidence of teacher characteristics influencing efficiency.

Competition is as discussed above the more important explanatory factor in the analysis. The nature of the competition and the laws under which private schools exists are not determined in the local school organization but rather by the market and by central government. The results on competition suggest that when continuing the analysis of school efficiency the role of competition needs to be studied more in depth. However, the local school is a natural starting point for improving efficiency. The role of teachers is weak in this study along with many others. This suggests that we have to extend the data with further explanatory variables before we can give appropriate policy recommendations for improving efficiency at the local level. One direction pointed out in the literature is organizational differences between schools. In 1991/92 the possibilities for Swedish municipalities to organize public education to fit local conditions were increased, and analyzing the organizational differences between efficient and inefficient schools closer could give additional suggestions on how to run a school successfully. The incentives for teachers and managers to be efficient is also emphasized in the literature. Organizational changes do not necessarily improve efficiency in themselves if the incentives for teachers and students to achieve high grades are not changed.

6 Summary

In this paper we estimate efficiency in Swedish public secondary schools in 1994/95. DEA models for school production are estimated with mean efficiency ranging from 0.81 to 0.91 depending on returns to scale assumptions and the inclusion of environmental variables. As variable inputs in the model are teachers and as outputs grades in mathematics, Swedish and English. These are regarded as the three most important subjects in Sweden and the grades are based on national tests taken by all students. As environmental variables are used the number of students in three socioeconomic groups based on parents' education and immigration. In the second part of the paper we analyze the estimated efficiency with statistical models to determine the influence on efficiency from private school competition and from schools employing teachers with different characteristics. The most important factor in explaining why units are more efficient is the competition from private schools. None of the teacher characteristics show a significant influence on efficiency.

References

Bergstrom, F. & Sandstrom, M. (2001), *Konkurrens Bildar Skola - En ESO-Rapport Om Friskolornas Betydelse För de Kommunala Skolorna*, Departementsserien 2001:12.

Bradley, S., Johnes, G. & Millington, J. (2001), 'The effect of competiton on the efficiency of secondary schools in england', *European Journal of Operational Research* **135**, 545–568.

Duncombe, W., Miner, J. & Ruggiero, J. (1997), 'Empirical evaluation of bureaucratic models of inefficiency', *Public Choice* **93**, 1–18.

Färe, R., Grosskopf, S., Försund, F., Hayes, K. & Heshmati, A. (1997), *Productivity and Quality in Swedish Schools*, Göteborg University, memorandum no 238.

Färe, R., Grosskopf, S. & Lovell, K. (1994), *Production Frontiers*, Cambridge University Press.

Färe, R., Grosskopf, S. & Weber, W. (1989), 'Measuring school district performance', *Public Finance Quarterly* **17**(4), 409–428.

Greene, W. (1997), *Econometric Analysis*, Prentice-Hall.

Grosskopf, S., Hayes, K., Taylor, L. & Weber, W. (1997), 'Budget-constrained frontier measures of fiscal equality and efficiency in schooling', *The Review of Economics and Statistics* **LXXIX**, 116–124.

Grosskopf, S., Hayes, K., Taylor, L. & Weber, W. (1999), 'Anticipating the consequences of school reform: A new use of DEA', *Management Science* **45**, 608–620.

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

Hanushek, E. (1996), Outcomes, costs, and incentives in schools, *in* E. Hanushek & D. Jorgensen, eds, 'Improving America's Schools - The Role of Incentives', National Academy Press, pp. 29–52.

Hedges, L. & Greenwald, R. (1996), Have times changed? the relation between school resources and student performance, *in* G. Burtless, ed., 'Does Money Matter? -

The Effects of School Resources on Student Performance and Adult Success', The Brookings Institution, pp. 74–92.

Heshmati, A. & Kumbhakar, S. (1997), 'Efficiency of the primary and secondary schools in sweden', *Scandinavian Journal of Educational Research* **41**(1), 33–51.

Hoxby, C. (2001), 'How school choice affects the achievement of public school students'. Koret Task Force meeting, September 2001, Hoover Institution, Stanford, CA.

Kirjavainen & Loikkanen (1998), 'Efficiency differences of finnish senior secondary schools: An application of data envelopment analysis and tobit analysis', *Economics of Education Review* **17**, 377–394.

Ljunggren, U. (1999), *En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan - med tillämpning i Stockholms stads grundskolor*, Stockholm School of Economics.

Lovell, K., Walters, L. & Wood, L. (1994), Stratified models of education production using modified DEA and regression analysis, *in* A. Charnes, W. Cooper, A. Lewin & L. Seiford, eds, 'Data Envelopment Analysis: Theory, Methodology and Applications', Kluwer Academic, pp. 329–351.

McCarty & Yaisawarng (1993), 'Technical efficiency in new jersey school districts', *The Measurement of Producer Efficency - Techniques and Applications by Fried, Lovell and Schmidt (eds)* .

Muñiz, M. (n.d.), 'Separating managerial inefficiency and external conditions in data envelopment analysis'. forthcoming in European Journal of Operational Research.

Olesen, O. & Petersen, N. (1995), 'Incorporating quality into data envelopment analysis: A stochastic dominance approach', *International journal of production economics* **39**, 117–135.

Skolverket (1999), *Samband mellan resurser och resultat - en studie av landets grundskolor med elever i årskurs9*, Rapport Nr 170, Skolverket.

Waldo, S. (2000), *Municipalities As Educational Producers - An Efficiency Approach*, Working Paper 2000:19 Department of Economics, Lund University.

Appendix

A Explaining Efficiency in CRS Models[11]

Dep variable:	Log of efficiency scores			
	MI		**MIII**	
	Coeff	**p-value**	**Coeff**	**p-value**
Const	-0.7545	0.0000	-0.2097	0.0026
Private schools	0.0108	0.0000	0.0099	0.0000
Continuity	0.0002	0.4919	0.0001	0.8887
Experience	0.0004	0.7657	0.0019	0.2957
Skill	0.2478	0.0009	0.0628	0.4553
Large city	0.0127	0.1321	0.0131	0.2300
Pure secondary	0.0295	0.0000	0.0003	0.9747
Parents' education	0.1433	0.0000	-	-
Immigrants	-0.0030	0.0000	-	-
Specification test				
Normality	LM = 15.02	0.0005	LM = 2.10	0.3501
No of units	847		725	

Table 7: Tobit Model Explaining the Efficiency Distribution in Model I and Model III (CRS)

[11] The models are corrected for heteroscedasticity in the same way as the models explaning VRS efficiency. In model I LR = 25.72 (p=0.0003) and in model III LR = 20.94 (p=0.0019).

B Figures

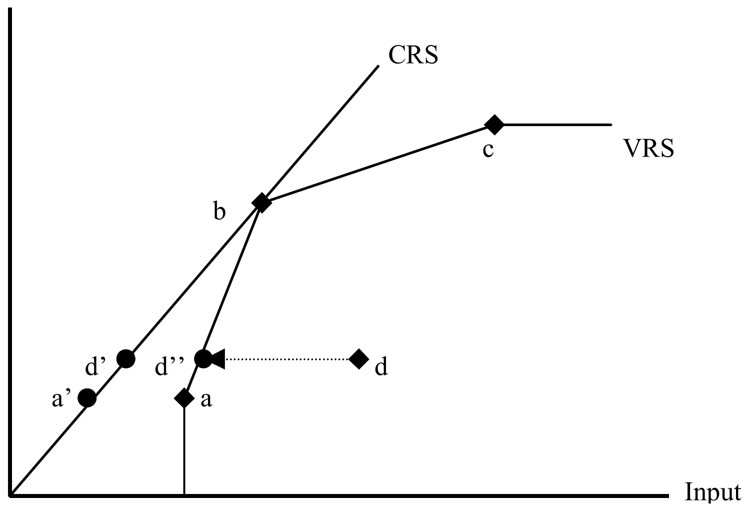

Figure 1: Technical Efficiency

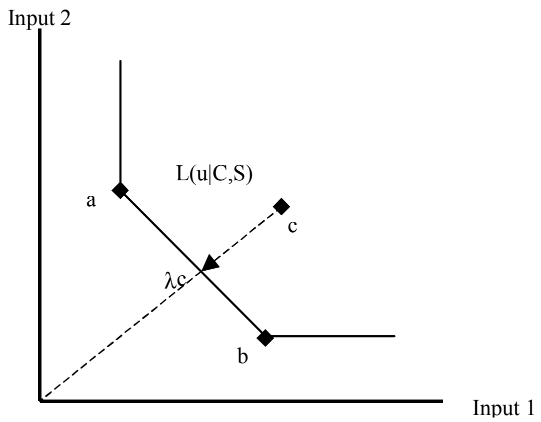

Figure 2: Technical Efficiency with two inputs