VATT-KESKUSTELUALOITTEITA
VATT-DISCUSSION PAPERS

91

SCHOOL RESOURCES AND STUDENT ACHIEVEMENT: EVIDENCE FROM FINNISH SENIOR SECONDARY SCHOOLS

Tanja Kirjavainen*
Heikki A. Loikkanen**

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 1995

* Researcher, Government Institute for Economic Research
Internet e-mail: tanja.kirjavainen@elisa.vatt.vatt.vatt.fi
** Professor of Local Public Economics, University of Tampere
Adr. Dept. of Local Government Studies, P. O. Box 607, 33101 Tampere, Finland.
Internet e-mail: heikki.loikkanen@uta.fi

Tanja Kirjavainen gratefully acknowledges the financial support from the Board of Education.

ISBN 951-561-129-6
ISSN 0788-5016

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Hämeentie 3, 00530 Helsinki, Finland

Painatuskeskus Pikapaino Opastinsilta
Helsinki, huhtikuu 1995

Kirjavainen, Tanja and Heikki A. Loikkanen: School Resources and Student Achievement: Evidence from Finnish Senior Secondary Schools. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 1995. 46 s. (C, ISSN 0788-5016; No 91) ISBN 951-561-129-6.

ABSTRACT: In this study we analyze how various factors (inputs) affect alternative achievement (output) measures in Finnish senior secondary schools. We have two output measures: results from national matriculation examinations in terms of scores and final report scores given by teachers of each school. As explanatory factors we have school related variables such as teachers' experience and education, class and school size. Alternatively we use pecuniary input measures like teachers' average salary or teaching expenditures per pupil. Pupil quality is controlled by lowest school report score at the time of admission to the school and by the education level of students' parents. Furthermore, we have dummy variables measuring degree of urbanization. In addition to linear and logistic models explaining average performance, we explain the standard deviation of performance.

According to our results parents' education level and the share of female students invariably have a positive effect on achievement regardless of output variable. In cities the achievement levels are lower than in less urban areas. Surprisingly, class size or teachers' experience and education either have no effect or it is in unexpected direction. School size sometimes has a positive effect on achievement as well as monetary measures of school resources. In addition to improving average performance in matriculation examinations, an increase in parents' education level reduces its standard deviation. In case of school reports, only admission level has a negative effect on variability of performance.

KEY WORDS: education, input-output analysis, production function.

Kirjavainen, Tanja and Heikki A. Loikkanen: School Resources and Student Achievement: Evidence from Finnish Senior Secondary Schools. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 1995. 46 s. (C, ISSN 0788-5016; No 91) ISBN 951-561-129-6.

TIIVISTELMÄ: Tässä työssä tutkitaan miten lukioiden resurssitekijät (panokset) selittävät niiden suoritustasoa (tuotosta). Suoritustasoa mitataan sekä ylioppilastutkintotodistuksen pistemäärällä että lukioiden päästötodistuksen keskiarvon perusteella. Suoritustasoa selittävinä tekijöinä ovat opettajien kokemusta ja koulutustasoa mittaavat muuttujat sekä luokka- ja koulukoko. Vaihtoehtoisesti näiden muuttujien tilalla käytetään opettajien keskimääräistä palkkaa ja opetusmenoja oppilasta kohti. Oppilasainesta kontrolloidaan lukioiden sisäänpääsyrajalla ja toisaalta oppilaiden vanhempien koulutustasolla. Lisäksi käytettävissä on asuinalueen kaupungistumisastetta kuvaavia muuttujia. Lineaaristen ja logististen keskimääräistä suoritustasoa selittävien mallien lisäksi samoilla muuttujilla selitetään suoritustason vaihtelua (standardipoikkeamaa).

Tutkimuksen tulosten mukaan vanhempien koulutustaso ja tyttöoppilaiden osuus vaikuttavat positiivisesti lukioiden suoritustasoon tuotosmuuttujasta riippumatta. Kaupungeissa suoritustaso on taajamia ja maaseutua alhaisempi. Yllättäen luokkakoolla ja opettajien kokemuksella ja koulutuksella ei ole vaikutusta suoritustasoon tai vaikutukset ovat suunnaltaan odotusten vastaisia. Rahallisilla resurssimitoilla ja koulukoolla on joissakin tapauksissa positiivinen vaikutus suoritustasoon. Vanhempien koulutustason kasvaessa suoritustason paranemisen ohella ylioppiskirjoitusmenetyksen vaihtelu (standardipoikkeama) pienenee. Koulun päästötodistusten vaihteluun puolestaan vaikuttaa vähentävästi sisäänpääsyrajan kasvu.

AVAINSANAT: Koulutus, panos-tuotos analyysi, tuotantofunktio

CONTENTS Page

1 INTRODUCTION

The purpose of this paper is to study the efficiency in the Finnish educational system. More specifically, we analyze the relation between resources and performance of Finnish senior secondary schools. In general, knowledge on how to promote education most effectively is of importance since education is viewed as the single most important factor behind economic progress. Furthermore, in countries like Finland, where high unemployment and budget deficits in the last few years have compelled to make budget cuts and public sector reforms, it is important to know the possible consequences of such measures.

Our topic is not a new one. Especially in the Anglo-Saxon countries, quite a few studies have tried to find out the relation between school inputs and outputs (c.f. Hanushek 1986). Also in Finland there has been a lot of educational research, but to our knowledge there has not been previous work using the educational production function approach we employ in this study.

In this study we construct regression models in which output measures are explained by variables related to schools' resources and quality and socioeconomic background of their pupils. In earlier studies the output variable of schools has most often been a test score measuring the students' educational achievement. In our case we shall use two alternative output measures, namely scores of national matriculation examinations and scores in final reports given autonomously by teachers of each school. As input variables we have teachers' experience and education as well as school and class size. In Finland senior secondary schools select their pupils by taking the best candidates from the pool of applicants on the basis of comprehensive school reports. Our measure of pupils' quality is the lowest score with which each school could be entered. Socioeconomic background of pupils is measured by their parents' level of education. Also sex distribution of students and degree of urbanization of area of school are controlled for. Finally, as an alternative to some non-pecuniary input variables, we also have information on educational expenditures per pupil and salaries per teacher.

In addition to linear and logistic models explaining average performance of schools, we also have models in which the variability of performance, measured by standard deviation, is explained basically with the same kind of variables as above. Namely, it is possible that a factor has no effect on average performance but affects its variation.

Alternatively, it may locally have similar or opposite effect on variation relative to its effect on average performance.

This paper is organized as follows. In chapter 2 we present a selective survey of studies in which regression type analysis has been employed to study the relationship between output and input variables of schools. In chapter 3 we present the framework of this study, available data and variables to be used in our analysis. In chapter 4 we present the results of our study. Chapter 5 offers a summary and some conclusions.

2 A SELECTIVE SURVEY OF INPUT-OUTPUT STUDIES

Peter Cuttance (1985) divides the studies on the effectiveness of schools into four different frameworks. They are the input-output framework, organizational framework, institutional framework and the exemplary school framework. This study can be characterized as an input-output study which aims at finding school and its environment related explanatory factors for variation of school achievement. Previous literature which uses this approach has been surveyed in Cohn and Geske (1990, ch. 7) and Hanushek (1986).[1]

In the applications of the input-output framework the output or achievement level of schools is typically measured by test scores of pupils. Differences in the achievement level of schools are explained by school related and environmental variables. Typical school related factors include variables at least to some extent controllable by the authorities or the schools themselves like the number and qualifications of teachers, school and class size and the quality of pupils. Environmental variables are beyond the direct control of schools and school authorities and they are typically measured by socio-economic background of pupils.

The measurement of school output is both difficult and controversial. Most often learning is viewed as the key output of schools. Accordingly, most studies use pupils' test scores as an output measure. When a broader view of the purpose of schools is taken, they are seen as providing skills for life. In other words schooling is thought of as an investment in human capital which is expected to increase people's productivity. The improvement of ones productivity is assumed to be reflected in ones salary. Thus one alternative output measure has been income at some time after school. The use of this kind of output variable is, however, more complicated than the use of test scores as the number of other factors affecting earnings after school increases.

In the following we shall survey previous input-output studies. We confine ourselves to statistical (parametric) studies which are similar to our own application. We will also rely to some extent on Eric Hanushek's (1986) review on earlier studies that have used the input-output framework.

---

[1] Studies using the exemplary school framework are considered in Purkey and Smith (1983). The British studies have also been reviewed by Reynolds (1992) and a corresponding work on US studies can be found in Levine (1992).

In his survey article Eric Hanushek (1986) reports the results of 147 studies where the relationship of inputs and outputs in public schools in United States is investigated. All the studies used either the individual, school or district level data (about half of studies used individual level data). The output was measured in 96 of the studies by using some standard achievement test. About 40 per cent of the studies concentrated on the variance inside single school district and in the rest 60 per cent the variation across school districts was considered. In two thirds of the studies high schools were investigated and the rest of the studies considered elementary schools. In all of the studies the social background of pupils was also controlled.

Table 1. The effect of class size, teachers' education, experience and salary on pupils' test scores in the United States.

		Statistically significant at 5 % level		Statistically insignificant at 5 % level			
	Number of Studies	+	-	Total	+	-	coefficient unknown
Teacher/ student-ratio	112	9	14	89	25	43	21
Teachers' education	106	6	5	95	26	32	37
Teachers' experience	109	33	7	69	32	22	15
Teachers' salaries	60	9	1	50	15	11	24
Expenditures/ pupil	65	13	3	49	25	13	11

Source: Hanushek, 1986 p. 1161.

In most of the studies covered by Hanushek's survey, the factors mentioned in table 1 did not have a statistically significant effect on students' achievement. In some studies the coefficient was statistically significant but had the "wrong" sign. Class size had expected negative effect on the achievement only in six per cent of the studies. Most important factor seemed to be the experience of teachers which had the expected positive effect on achievement in 30 per cent of the studies. Teachers' salaries affected achievement positively in about 15 percent of the studies and expenditures per pupil had a positive effect in 20 per cent of the studies.

On the basis of his survey Hanushek concludes that there is no strong evidence on a positive relationship between schools' resources and pupils' achievement measured by test scores. As for factors outside the control of schools, sosioeconomic background is a key explanatory variable.

The discussion about factors affecting school performance has continued also after Hanushek's survey article. In recent years some studies using input-output approach have been published (Ehrenberg & Brewer (1994), Sander (1993), Dolan & Smith (1987)). These studies fulfill the basic requirements of Hanushek's survey article in the sense that in all of them the socioeconomic status of pupils has been controlled. Two of these studies use in addition to test scores also other school related output variables.

The results of the more recent studies are also somewhat ambiguous. According to Ehrenberg and Brewer (1994) the resources of the schools and the variables that represent the characteristics of teachers rarely predict the probability of student dropping out of high school or the change in the achievement if the student did not drop out. However, in some cases these factors have an influence. Sander (1993) reported that class size had the expected negative effect on test achievement, graduation rate and the rate of continuing studies in college. Teachers' salaries had a positive effect on students' achievement in some cases. Dolan and Smith (1987) studied the effect of expenditures on student achievement in elementary schools. Also their results are mixed. Sometimes expenditures had a positive effect and sometimes they did not.

Similar, mixed results have also been reached using international data by the Association for the Evaluation of Education and Achievement (IEA). The study done in 1975 by Simmons covered 258 000 pupils, 5 900 teachers and 9 700 schools from 23 countries. According to the results school related factors had very limited power in explaining achievement differences when the role of other factors was controlled. The latest study that used IEA data (Lewis and Seidman, 1994) examined factors that influence the mathematics learning of eight grade students. The results showed that the time used for teaching mathematics was the only statistically significant explanatory factor. Teachers' experience and education as well as school size proved to be statistically insignificant in explaining the test performance.

There are also studies concerning developing countries. According to a survey on these studies (Alexander and Simmons, 1978) the results are mixed. School factors such as class and school size sometimes have the expected sign but in a number of other studies they have no statistically significant effect or they have an unexpected effect. On the basis of studies concerning developing countries Fuller (1978) concludes that in the Third World school related factors are more important determinants of performance than in developed countries when the socioeconomic background of pupils has been controlled.

In Nordic countries the type of input-output studies surveyed by Hanushek (1986) are either rare or non-existent. In Sweden there is a study comparing the performance of primary school pupils and school costs in 29 municipalities (Finansdepartementet, 1986). No connection was found between the costs per pupil and the achievement indicators, except for the fact that there was a positive connection between expenditures of school materials, audiovisual equipment and library, and how students get along. In Finland a study closest to this approach is Sandberg (1992) where the author evaluates the performance of primary schools. It is however, not in effect an input-output study as the performance indicator is a mixture of input and output factors.[2]

Finally, we note that the studies surveyed have all measured performance of schools by test scores or other indicators (drop out rates, continuation into college etc.) measured either during or at the end of school. There are also studies in which the output measure is related to a later phase in life, namely to labor market performance measured by earnings. In a study by Card and Krueger (1992) the weekly earnings of males born in the US during 1920-49 were explained i.e. with variables related to their schools. The data was micro data on individuals but the information of schools was based on state averages. According to the results educational investments have a higher return for individuals who have been in schools with lower pupil/teacher ratio, and the teachers had relatively higher wages and educational levels. Also the role of parents' income and education level was studied on the basis of state level (average) variables. Here, unlike in most studies explaining variation of test scores, they did not get statistically significant coefficients in the estimated regression models.

To conclude, according to our selective survey the empirical research results of statistical input-output analyses of the performance of schools are mixed. Factors like the experience of teachers, education of teachers, class size and school expenditure per pupil, which usually are supposed to affect the performance of schools may or may not have a

[2] Above we have not included in our survey studies which have applied non-statistical methods. We note that there are Nordic applications of Data Envelopment Analysis (Bonesrønning and Rattsø 1995, Kirjavainen and Loikkanen, 1993).

clear relationship to output variables. Next, we shall turn our attention to our own analysis in which we conduct a regression analysis trying to find out whether the same kind of input variables used in previous studies explain the performance of Finnish senior secondary schools.

3 MODELS, DATA AND VARIABLES

3.1 Models

The general form of the educational production function can be described by the following equation (1), which was introduced by Hanushek (1971, 1979):

$$A_{it} = f^* (B_i^{(t-t^*)}, P_i^{(t-t^*)}, S_i^{(t-t^*)}, I_i, A_{it^*}) \quad (1)$$

where

A_{it} = student i's achievement at time t
$B_i^{(t-t^*)}$ = vector of family background influences over the period t* to t
$P_i^{(t-t^*)}$ = vector of influences of peers over the period t* to t
$S_i^{(t-t^*)}$ = vector of school inputs of *ith* student over the period t* to t
I_i = vector of innate abilities of *ith* student
A_{it^*} = outcomes of the *ith* student in earlier period

This formulation is also called the "value added" model as opposed to the cumulative model where all the inputs that the student has received during her/his life are assumed to affect her/his performance. This formulation is more commonly used because it reduces the data requirements. It allows to evaluate the educational services the student has received in the school by taking also into account the influences of student's innate abilities, former outcomes, family background and peers.

In the applications of the value added model the data may consist of information on individuals or it may be aggregated to for example school level. In our case the units of observation are schools and the performance indicators measure the average achievement of pupils in each school. From a statistical or econometric view-point, the input-output analyses have included applications from correlation analyses to more sophisticated methods in which the effects of inputs to a multidimensional output vector have been estimated. Here, we shall confine ourselves to single-equation regression models.

There will be of three types of models and as a fourth issue we shall discuss pooling of data. First, like in most other previous studies, we estimate simple linear regression models. Second, we recognize that both of our alternative dependent variables are in effect bounded both from below and above and estimate logistic regressions, i.e. models in which the dependent variable is transformed in such a way that the predictions of the model remain in the natural bounds of the dependent variables. Third, we shall study how variability of performance is affected by different factors. For this purpose, we adopt the procedure applied by Brown and Saks (1975) and estimate models in which the standard deviation of performance indicators are regressed with our independent variables. Fourth, we note that we have performance indicators for the same schools from two points of time, and related to this situation we shall consider the alternatives of estimating separate models for the two years.

3.1.1 Linear specification

By denoting a dependent variable (performance indicator) by Y_j for school j and the independent variables i (i = 1,...,m) by X_{ij} the linear models to be estimated in this study are of the following form

$$w_j Y_j = \alpha + \Sigma \beta_i w_j X_{ij} + e_j \qquad (2)$$

where

α = constant term
β_i = coefficient of variable i
w_j = weight of observation j
e_j = the error term

Referring to equation (1), there are variables which should cover the relevant school period from t* to t. In our case, the Finnish senior secondary school lasts for three years. Most of our school related resource variables are averages over the three year period, whereas the socioeconomic and location of school variables refer to a single year. The students outcomes in earlier periods is measured indirectly by the schools' admission level. A more detailed description of the variables to be used will be given in the sequel.

As for equation (2) we shall first estimate linear (in parameters) models by ordinary least squares (OLS) with weights (w_j). The weights are the number of students taking matriculation examination in each school as e.g. in Sander's (1993) study. As a result of weighting, large schools will get a greater weight. We shall also discuss results with no weighting without presenting these results, however, in this paper.[3]

3.1.2 Logistic specification

A problem with the linear model in equation (2) is that it cannot be a correct specification as its predictions will not remain within the range of dependent variables with more extreme values of independent variables.[4] One way to allow for non-linearity in the effects of independent variables and to quarantee that the predictions of the model within the relevant ranges is to apply a logistic transformation. For this purpose let us denote the dependent variable as above by Y_j for schools j and its range by $[Y_{min}, Y_{max}]$. Then, define variable P_j as an odds ratio

$$P_j = (Y_j - Y_{min}) / (Y_{max} - Y_{min}) \quad (3)$$

Note that if school j has the best possible score Y_{max}, then $P_j = 1$, and if it has the lowest possible mark Y_{min}, then $P_j = 0$. Next, let us define a log odds type variable S_j (so called logit) as

$$S_j = ln[P_j / (1-P_j)] \quad (4)$$

and assume that S_j is a linear function of independent variables X_{ij} ($i = 1,...,m$) i.e.

$$S_j = ln[P_j / (1-P_j)] = \alpha_i + \Sigma \beta_i X_{ij} \quad (5)$$

from which P_j can be solved as

$$P_j = 1 / [1+exp(-(\alpha_i + \Sigma \beta_i X_{ij}))] \quad (6)$$

Equation (6) is the so called logit model which can be estimated by Maximum Likelihood Estimation (MLE) method if the dependent variable is a dichotomous (one or zero) type variable such the Pj can be interpreted as a probability. In our case we shall scale school scores to [0,1] interval using (3) and then we shall estimate (5) by OLS and use (3) to solve for an estimate of Y_j for school j as follows

$$Y_j = Y_{min} + P_j (Y_{max} - Y_{min}) \quad (7)$$

[3] The results of these estimations are available upon request.

[4] Recall that the matriculation examination scores at individual level have a range from 1 (fail) to six (laudatur). Respectively, in the final reports given by school teachers the range is from four (fail) to ten (best mark).

In our application we have two alternative dependent variables with which utilize equations (3)-(7). In the case of national matriculation examinations $Y_{min} = 1$ and $Y_{max} = 6$. As for final school reports, $Y_{min} = 4$ and $Y_{max} = 10$, respectively. In order to point out how in this setting independent variables X_{ij} affect the dependent variable, we refer to equation (5)-(6). If a variable gets a positive (negative) and statistically significant coefficient when estimating (5), then it will affect positively (negatively) to log odds of P_j and thus also to P_j itself and according to (7) positively (negatively) also to Y_j. As the specification is non-linear, the marginal effects of independent variables are not constant. By denoting the RHS of (6) by Z, we can say that as Z increases, then P_j increases. Graphically their relation is of S (sigmoid) type such that as $Z = -$ infinity, $P_j = 0$ and $Y_j = Y_{min}$. Respectively, as $Z = +$ infinity, $P_j = 1$ and $Y_j = Y_{max}$.

The above presented equation (5) is estimated both without weighting and with weights (w_j) as equation (2). We will present the results of weighted regressions in this paper. The results of non-weighted regressions are available upon request.

3.1.3 Model explaining variability of performance

The above linear and logistic models aim at explaining the relation between average performance and average resources of senior secondary schools. As pointed out by Brown and Saks (1975) input variables may also affect in addition to mean performance to their distribution. As a special case it is possible that mean performance is not affected by some variables at all but the distribution is affected. It is also possible that some variable has an opposite effect on mean and standard deviation of performance. To test for this kind of possibilities, we also run weighted OLS regressions of the following type

$$t_j D_j = \alpha + \Sigma t_j \beta_i X_{ij} + e_j \qquad (8)$$

where D_j is the standard deviation of the performance indicator for school j, t's are weights and X_{ij}'s are values of independent variables i (i = 1,...,m).

3.1.4 On models with pooled data

Finally, we note that we could also have made other choices with our data and model type. As we have two obsevations (1991 and 1992) on the same school we could have treated them as a panel and modeled it e.g. as a fixed effects model. As the panel would have been short we omitted this possibility. Instead we pooled the observations treating each school in each year as a separate observation and estimated models with a year-dummy. Results of estimating linear (c.f. equation (2)) models with pooled data will be in Appendix 4. Results of logistic specification (c.f. equation (5)) using pooled data are in Appendix 5 and the results of explaining the variability of school performance (c.f. equation 8)) with pooled data are in Appendix 6. They all will only be commented shortly at the end of chapter 5.

3.2 Data

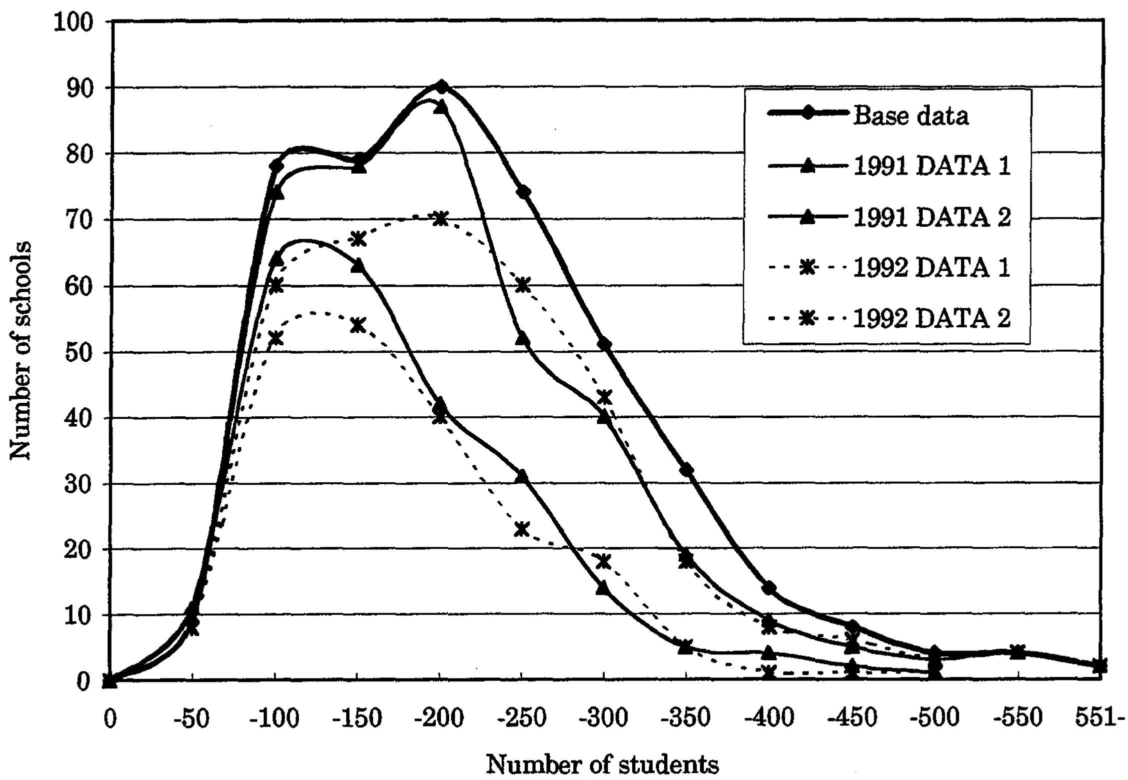
The data used in this study is cross-sectional and aggregated to the school level. The data has been constructed from several registers and a survey. Information on pupils' grades in matriculation examinations and grades in final school reports are from a register of the matriculation examination committee held at VTKK (state owned computer and information system company). Information on teaching hours, education and experience of teachers, number of pupils are from so called PELU-register also held by VTKK. The expenditure information is from a register held by the University of Tampere and the Board of Education. Admission level information has been obtained by our own survey directed to schools. Variables related to the socioeconomic background of students (the average education level of students' parents) are from the census of Statistics Finland.

The regression models of the study will be estimated with four different data in which the number of observations varies. They can be classified both by the year of graduation (1991 and 1992) and on the basis of whether cost information can be attributed to individual schools. In the latter case there are two data alternatives. The first one (DATA 1) contains all schools on which we have survey information concerning admission level (384 schools in 1991 and 350 schools in 1992). The second one (DATA 2) contains only such schools which are the only ones in their municipality (239 schools in 1991 and 209 schools in 1992) except for the private schools. In the first case, if several senior secondary schools are located in the same municipality, the costs cannot be attributed to individual schools, rather they are municipality averages except for private schools for which school specific cost information is available in all cases.

Especially in the case with school specific cost information small schools are overrepresented in the respective data set. This can also be seen from figure 1 in which the school size distributions of the four data alternatives are displayed together with the distribution concerning all schools. One also notes that the data for 1991 (part of the data is based on municipality averages) is closest to the size distribution of all schools.

The two data used in estimations have their advantages and disadvantages. On the one hand, the first data set (DATA 1) is larger and in that way more representative. On the other hand, the second data set (DATA 2) even though it is smaller, some of its variables are more precise.

Figure 1. Frequency distribution of school size by number of pupils in our four data alternatives compared to all schools.



In this study we shall present estimation results of regression models in which the dependent variables will be an indicator of achievement in matriculation examinations and the scores given by teachers at the end of the school. Separate models will be estimated for two different years, 1991 and 1992. Because the senior secondary school is a three year school those who graduated in 1991 started their studies in Fall 1988 and those who graduated in 1992 started their studies in Fall 1989. The achievement scores (school averages) will be explained by respective resource and cost variables which are averages over three proceeding school years except for some resource variables (teachers' education and experience, and school and class size) that are only two year averages

(1989-91) for those who graduated in 1991. Naturally, the admission level variables refer to either Fall 1988 or Fall 1989. In general the independent variables are similar to those used in earlier studies.

3.3 Variables

The output or performance of schools will be measured by two different variables. The first independent variable is students' overall score in matriculation examination in spring 1991 and in spring 1992. The score is derived by giving six points for laudatur, five for magna cum laude approbatur, four for cum laude approbatur, three for lubenter, two for approbatur and one for improbatur (failed) and finally by calculating the school average on the basis of these scores of individual pupils. The average score was 4.23 in 1991 and 4.22 in 1992 in our DATA 1.

The second dependent variable is the average score in the students' final reports. It consists of subjects that may be characterized as theoretical i.e. mathematics, languages and humanities and natural sciences. It is an average of each of the subjects and it is given by teachers on the basis of students' performance at school. The scale of the scores ranges from 4 to 10 where number four means failed and number 10 excellent performance. The average score was 7.86 in DATA 1 in 1991 and in 1992.

This study in other words considers two different output measures which are either a more 'objective' matriculation examination score or a more 'subjective' final report score given by teachers. These two measures are different from each other. The correlation of these two output variables was 0.54 in 1991 and 0.56 in 1992 in DATA 1. It is not in other words as high as one would expect.

Following the variable categorization of Hanushek (c.f. equation (1)) we have the following input factors in our models. The students' outcomes in earlier periods are measured by the schools' admission level. Family background is controlled by a variable measuring the educational level of students' parents. One variable that in some sense reflects student's innate abilities is sex. It is measured in this study by the percentage share of female students in the school. Resource input variables are measured by teachers education and experience, class and school size. In some alternative models we included instead of the teachers' experience and education teachers' average salaries or teaching expenditures. The dummy variables related to population density of the schools' environment may be said to reflect among other things the influence of peers.

Pupils to senior secondary schools are chosen by a "cream skimming" procedure, i.e. by choosing the best applicants on the basis of their comprehensive school reports (graded using a scale from four to ten). Here, the quality of pupils is controlled by the lowest grade with which the school could be entered (*ADMISSION LEVEL*). As the school lasts for three years, the lowest grades are from the years 1988 and 1989. Only information on this cut off point could be obtained by our survey, although the average grade of admitted pupils would have been a better measure.

The socioeconomic background of pupils is measured by the average education level of students' parents (*PARENTS' EDUCATION*). The variable is constructed at the individual level by giving points according to degree.[5] Here we use the school level averages. As the parents' average educational level increases, the performance of pupils is expected to improve. This is because more educated (and better earning) parents may provide a better resources and environment for learning at home that also helps learning and performance at school.

In some input-output studies the gender distribution is also taken into account i.e. student's sex is assumed to have an influence on performance. In this study the distribution of sex (*PERCENTAGE FEMALE*) in the school is measured by the percentage share of female students passing the matriculation examination. It is based either on the results of 1991 or 1992 examination results depending on the dependent variable. We have no prior assumption about the sign of this variable.

Teachers are the most important single input of schools. Their education and experience are expected to affect performance of schools. In this study the education level of teachers is measured by the share of teachers having at least a masters degree (*TEACHERS' EDUCATION*). It is either an average share for the years 1989-91 or the years 1989-92. Higher education level of teachers is expected to have a positive effect on the performance of schools.

---

[5] The degrees are given the following points: 1,5 = no other degrees than comprehensive school diploma. 3 = lowest vocational degree (approximately 10-11 years of schooling). 4 = medium vocational degree (approximately 12 years of schooling). 5 = highest vocational degree, not a university degree (13-14 years of schooling). 6 = bachelor's degree. 7 = master's degree. 8 = post graduate degree.

A variable measuring teachers' experience (*TEACHERS' EXPERIENCE*) is indirectly based on the number years they have taught. Namely, lacking direct information we utilize salary information because teachers get seniority bonuses according to a certain time schedule. We use the information on teacher specific bonuses to calculate a proxy variable of teacher experience as an average over the periods 1989-91 and 1989-1992.

In some model variants teachers' average salary (*AVERAGE TEACHERS' SALARY*) is substituted for variables measuring teachers' education and experience. Depending on the year of the dependent variable it is either an average over 1988-91 or over 1989-92. Salaries from different years have been deflated by a salary index for public sector employees to 1991 level. In addition to that we have eliminated regional variation of salaries to the extent that it is based on regional pay schedules in order to get a variable which reflects teachers' education and experience (Appendix 1). We also divided the variable by 100 000 for matters of convenience. Teachers' salary is expected to have a positive effect on pupils' performance.

One of the variables which is most often included in studies on school performance is class size (*CLASS SIZE*). Because average class size is hard to measure directly in the case of senior secondary schools, instead of a direct measure we use a proxy. It is calculated by first dividing the number of teaching hours in the school by 20 (our estimate of the average teaching hours of each teacher per week) and then dividing the obtained number with the number of pupils. This variable is also an average over two or three years depending on the dependent variable and it is divided by 1 000 for coefficients to be of proper size. Small class sizes are expected to have a positive effect on performance.

Schools' teaching expenditure per pupil (*AVERAGE TEACHING EXPENDITURE/PUPIL*) is an alternative variable to teachers' education, experience and class size variables, i.e. a summary measure of resources devoted to teaching. Here as well we have eliminated regional variation in teachers' salaries and deflated them by the salary index for public sector employees to 1991 level. We shall also divide the figure by 10 000 for the coefficients to be of proper size. The variables to be used are averages over the relevant three year periods. In municipalities with only one senior secondary school, the variable is school specific. In municipalities with several schools, it is an average figure except for private schools for which it is school specific. Greater teaching expenditures per pupil are expected to affect performance positively.

School size is also one factor that may influence the performance of schools. It is often claimed that learning is more effective in small schools because, for example, the atmosphere is more intimate and teachers know each student personally. On the other hand, big schools which offer a larger variety of subjects may attract students that are more willing to work hard. This may also lead to higher achievement. In this study school size variable (*SCHOOL SIZE*) is also included in the models. It is measured by the number of students in the school. We have no prior expectation concerning the sign of the school size variable at this point.

The environment of school may also be one factor that has to be controlled for in the models. It is often claimed that schools in the rural area have a different kind of working environment compared to schools in more urban areas. In this study we have constructed dummy variables for the environment of school. It divides the municipalities into three different groups and it is based on the population density classification of Statistics Finland. In the first group, which is also the group of comparison, there are those municipalities which can be characterized as rural areas. The second group consists of municipalities that are densely populated but not urban (*DENSELY POPULATED AREA=1*). In the third group we have urban municipalities (*URBAN AREA=1*).

When the variability of performance is considered the independent variable is either the standard deviation of matriculation examination score or the standard deviation of final report score in each school. The standard deviation of matriculation examination score varied from school to school in 1991 from 0.64 to 1.82 points and from 0.60 to 1.90 points in 1992 in DATA 1. The variability of final report scores from school to school ranged in 1991 from 0.45 to 1.30 points and in 1992 from 0.49 to 1.24 points. We have no prior assumptions about the effects of independent variables on the variability of performance. Summary statistics of all the variables used are in Appendix 2.

4 ESTIMATION RESULTS

We shall first (section 4.1) present results of weighted least square estimation of linear production function. In section 4.2 we present some results of logistic specification, and in section 4.3 we have some results of running regressions with schools' standard deviation of performance indicator as the dependent variable. Finally, in section 4.4 we comment some results based on models using pooled data.

4.1 Linear specification

4.1.1 Matriculation examination score as a dependent variable

In table 2 we have results for the matriculation examination scores as the dependent variable. They are based on data (DATA 1) which mainly consists of school specific information except for some schools for which expenditure data are averages over all schools in the municipality, and the smaller data (DATA 2) which consists of school specific data for all independent variables. On the basis of estimated linear regressions we first note that the overall explanatory power of the models were rather modest. The R-squares[6] were at the lowest 0.12 and at the highest 0.29, i.e. at best about 30 per cent of the variation of average matriculation examination score was attributable to the explanatory factors used. In the following we shall comment the signs and statistical significance of variables in different models using the 5 per cent significance level as a benchmark.

In the first model (Model 1) in table 2 the average matriculation examination score is explained by admission level, parents' education, percentage share of female students, teachers' education and experience, class size, school size and area dummy-variables based on population density. The admission level, i.e. the lowest score in primary school reports among the admitted pupils, gets an expected positive sign in 1992. In 1991 the sign is positive but statistically insignificant. The socioeconomic variable measuring the educational level of the parents gets an expected positive and statistically significant coefficient in both years. The percentage share of female students has also a positive and

[6] Because the use of weighted least squares is only a computational device it is not of interest to calculate the R-squares from transformed variables. In this paper the R-squares are calculated using the measure proposed by Greene (1993, p. 363) in which the residuals are from the original models.

statistically significant coefficient in both years. In other words, the higher the proportion of girls the higher the examination scores. The variable related to teachers' experience is insignificant. The variable measuring teachers' educational level has surprisingly a negative and statistically significant coefficient. This would imply that the greater the share of teachers with master's degree, the worse the performance of pupils in the matriculation examination.

Our measure of class size and school size are both statistically insignificant. The area where school is situated has an effect on the average matriculation examination score. It was measured by dummy variables with rural area schools as the standard of comparison as opposed to more densely populated municipalities or urban areas. Schools that are situated in urban areas seem to have lower average scores than schools situated in densely populated or rural areas. Thus, according to model 1 students' socioeconomic background, quality of students and percentage share of female students have a statistically significant positive effect on student performance. In addition to that schools that are situated in rural areas seem to succeed better than their counterparts in more densely populated or urban areas. School related variables such as teachers' experience and education, and class size are either insignificant or affect achievement in unexpected ways.

In model 2 teachers' average salary has been substituted for teachers' education and experience. This salary variable gets an expected positive coefficient which however is statistically insignificant in both years. Variables measuring parents' educational level, admission level and percentage share of female students have all positive and significant coefficients in both years. The proxy for class size remains insignificant as well as the school size variable. The effect of school environment i.e. the coefficients of area-dummies behave similarly as in model 1 implying poorer results in urban areas.

In model 3 teaching expenditures per pupil variable is included in the model instead of teachers' education and experience, and class size. This variable gets a positive and statistically significant coefficient in 1991 but not in 1992. Other variables behave in the same manner as in model 2.

Table 2. Estimation results of weighted linear regression models using the average scores in matriculation examination in 1991 and in 1992 as the dependent variables.

	Average score in matriculation examination 1991					**Average score in matriculation examination 1992**				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**	**Model 1**	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.046521 (1.608)	0.042235 (1.449)	0.064299 (1.544)	0.039797 (1.368)	0.066217 (1.597)	0.057917** (1.955)	0.056703** (1.904)	0.012161*** (2.794)	0.052809* (1.766)	0.012485*** (2.860)
PARENTS' EDUCATION	0.005133*** (12.44)	0.005171*** (12.58)	0.004365*** (7.742)	0.005092*** (12.40)	0.004373*** (7.835)	0.004858*** (12.71)	0.004977*** (13.09)	0.003944*** (6.176)	0.004893*** (12.84)	0.003871*** (6.096)
PERCENTAGE FEMALE	0.85042*** (4.367)	0.88165*** (4.531)	0.44013 (1.558)	0.86104*** (4.421)	0.40057 (1.419)	0.66416*** (3.446)	0.67054*** (3.467)	0.57440** (2.195)	0.67974*** (3.519)	0.56652** (2.159)
TEACHERS' EXPERIENCE	0.0028781 (0.08190)					0.029318 (0.7705)				
TEACHERS' EDUCATION	-0.35062** (-2.165)					-0.38244** (-2.190)				
AVERAGE TEACHERS' SALARIES		0.10447* (1.695)	0.17224** (2.470)				0.083416 (1.332)	0.12202* (1.822)		
CLASS SIZE	-0.0043110 (-0.7617)	-0.009527* (-1.679)	-0.0054838 (-0.7811)			-0.0054404 (-0.8957)	-0.0098275* (-1.660)	-0.0032619 (-0.4643)		
TEACHING EXPENDI-TURES/PUPIL				0.10825** (2.033)	0.13995** (2.431)				0.060103 (1.131)	0.057965 (1.083)
SCHOOL SIZE	-0.053793 (-0.3148)	0.025943 (0.1544)	-0.17328 (-0.6966)	0.13117 (0.7315)	0.10855 (0.4042)	0.047085 (0.2556)	0.09937 (0.5444)	-0.13620 (-0.5041)	0.15049 (0.7970)	-0.0009539 (-0.003395)
URBAN AREA	-0.13437** (-2.552)	-0.15588*** (-2.945)	-0.12137 (-1.467)	-0.14063 (-2.697)	-0.12098 (-1.466)	-0.14390** (-2.533)	-0.15788*** (-2.757)	-0.085482 (-1.013)	-0.14929*** (-2.621)	-0.075827 (-0.8987)
DENSELY POPULATED AREA	-0.017171 (-0.3270)	-0.022266 (-0.4239)	-0.049846 (-0.8648)	-0.0012652 (-0.02375)	-0.018474 (-0.3194)	0.018986 (0.3214)	0.019256 (0.3255)	0.028680 (0.4628)	0.024385 (0.4030)	0.041760 (0.6635)
CONSTANT	2.2987*** (8.049)	1.8965*** (7.305)	2.0523*** (5.705)	1.7515*** (7.166)	1.9933*** (5.669)	2.3114*** (8.058)	1.9744*** (7.360)	1.7158*** (4.889)	1.8745*** (7.582)	1.7698*** (5.178)
$\tilde{R}^2$	0.211	0.212	0.120	0.230	0.149	0.290	0.292	0.192	0.289	0.182
N	384	384	239	384	239	350	350	205	350	205

***$p < 0.01$ **$p < 0.05$ *$p < 0.10$

Model 2 and model 3 were also estimated using DATA 2 which was smaller but more precise. Comparison of results in table 2 indicates that they do not change drastically as the data set decreases by the exclusion of schools for which we were only able to construct expenditure variables as municipality level averages. When the smaller data set is used, the percentage share of female students as well as class size loose their statistical significance. Neither do the area-dummies any longer have an effect on schools' performance.

Model 3 was also estimated using the smaller data. Also in this case the changes were small. In the smaller sample the percentage share of female students does not have an effect on performance in 1991 in the same manner as in model 2. The area-dummies controlling the environment of schools loose their significance, too.

The above models were also estimated without weighting. From a statistical view-point the results of unweighted models were worse than those reported above. School related resource variables and quality of pupils (admission) had less explanatory power when observations were not weighted by the number of pupils taking matriculation examination.

4.1.2 Final report score as a dependent variable

We also tested whether there were differences in the results depending on the performance measure used. For that purpose we estimated the models using the final report score as the dependent variable. This performance measure differs from the matriculation examination score in the sense that it is given by teachers. In other words, it reflects teachers' own view and tests of their students' learning whereas scores in the national matriculation examination are marked by persons outside the school. In table 4 we report the estimation results for average final report scores as independent variable. As we shall see, there were some differences in the estimation results when the dependent variable was changed to the average final report scores.

The explanatory power of these models was little lower than in the case of matriculation examination scores. The R-squares were at the lowest 0,12 and at the highest 0,22. In other words, at best only 22 percent of the variation in average final report scores was attributable to the factors used in this study.

When a closer look is taken the results show that in model 1 the admission level gets a positive and statistically significant coefficient in both years. The same also applies to the parents' education and the percentage share of girls which had the same sign also in the case of matriculation examination scores. It means that the higher the admission level, the educational level of parents and the percentage share of girls the higher scores teachers are giving. The resource level of schools does not affect the scores given by teachers. In other words, teachers' education and experience as well as class size get statistically insignificant coefficients in both years. School size gets a positive and statistically significant coefficient in both years. It means that teachers in larger schools give higher marks to their students contrary to the matriculation examination scores in case of which the school size did not affect performance.

The schools environment affects to the final report scores. Teachers who work in schools situated in rural areas give higher marks than their colleagues in urban and densely populated areas. Unlike with matriculation examination scores, there is also a statistically significant difference in marks between schools in urban areas and schools in densely populated areas relative to schools in rural areas. The marks are lowest in the urban areas here, too.

In model 2 teachers' average salary were substituted for the variables measuring teachers' education and experience. In other respects the model was the same as model 1. The results did not change as compared to model 1. Teachers' average salary was not a statistically significant explanatory factor. In other words, it does not affect the students' marks.

In model 3 variables measuring teachers' education and experience and class size were replaced by teaching expenditures per pupil. This variable had a positive and statistically significant coefficient in 1991 but not in 1992. Otherwise the model was similar to model 1 and model 2.

Models 2 and 3 were also estimated using the smaller sample (DATA 2). The results did not change very much. The only difference was that school size did not appear to be statistically significant in either of the models when the smaller sample was used, and this result applied to both years. When the results were compared to those using matriculation examination score as the dependent variable, they turned out to be similar to those of using the larger sample, i.e. input variables controllable by the schools

Table 3. Estimation results of weighted linear regression models using average final report scores in 1991 and 1992 as dependent variables.

	Average final report score in 1991					**Average final report score in 1992**				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**	**Model 1**	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.065780*** (3.524)	0.06280*** (3.356)	0.076460*** (2.878)	0.061184*** (3.283)	0.075335*** (2.855)	0.080857*** (4.342)	0.080109*** (4.294)	0.069733*** (2.694)	0.077461*** (4.151)	0.070664*** (2.729)
PARENTS' EDUCATION	0.002226*** (8.357)	0.002232*** (8.454)	0.002312*** (6.425)	0.002192*** (8.332)	0.002278*** (6.414)	0.001776*** (7.387)	0.001805*** (7.575)	0.002683*** (7.063)	0.001769*** (7.440)	0.002649*** (7.033)
PERCENTAGE FEMALE	0.6493*** (5.167)	0.65782*** (5.264)	0.57010*** (3.162)	0.64304*** (5.153)	0.55087*** (3.066)	0.53208*** (4.390)	0.53361*** (4.404)	0.72018*** (4.627)	0.53708*** (4.455)	0.71623*** (4.604)
TEACHERS' EXPERIENCE	0.0061602 (0.2716)					-0.00028568 (-0.01194)				
TEACHERS' EDUCATION	-0.088859 (-0.8501)					-0.088450 (-0.8054)				
AVERAGE TEACHERS' SALARIES		0.59293 (1.499)	0.67913 (1.526)				0.22401 (0.5713)	0.57920 (1.454)		
CLASS SIZE	-0.0018787 (-0.5144)	-0.0038076 (-1.045)	-0.0055170 (-1.231)			-0.0019775 (-0.5178)	-0.0033306 (-0.8981)	-0.0013455 (-0.3220)		
TEACHING EXPENDI-TURES/PUPIL				0.68405** (2.005)	0.73418** (2.004)				0.33230 (1.002)	0.33820 (1.066)
SCHOOL SIZE	0.22733** (2.061)	0.24853** (2.303)	0.13824 (0.8708)	0.32137*** (2.797)	0.26278 (1.538)	0.29802*** (2.573)	0.31036*** (2.714)	-0.0067965 (-0.4229)	0.33846*** (2.872)	0.007 (0.04201)
URBAN AREA	-0.22735*** (-6.691)	-0.23592*** (-6.940)	-0.18082*** (-3.425)	-0.22406*** (-6.708)	-0.18374*** (-3.500)	-0.23241*** (-6.506)	-0.23694*** (-6.606)	-0.20490*** (-4.082)	-0.23279*** (-6.547)	-0.19996*** (-3.996)
DENSELY POPULATED AREA	-0.11085*** (-3.271)	-0.1132*** (-3.358)	-0.088922** (-2.417)	-0.096999*** (-2.842)	-0.077240** (-2.098)	-0.10285*** (-2.769)	-0.10353*** (-2.794)	-0.073592** (-1.997)	-0.09820*** (-2.600)	-0.065435* (-1.753)
CONSTANT	6.5210*** (35.38)	6.3970*** (38.37)	6.3687*** (27.74)	6.3323*** (40.43)	6.2698*** (28.02)	6.6042*** (36.61)	6.5057*** (38.71)	6.1499*** (29.46)	6.4449*** (41.77)	6.1701*** (30.44)
$\tilde{R}^2$	0.197	0.194	0.124	0.217	0.151	0.228	0.228	0.202	0.227	0.199
N	384	384	239	384	239	350	350	205	350	205

***$p < 0.01$ **$p < 0.05$ * $p < 0.10$

remained insignificant whereas variables controlling for students' background and living area have statistically significant effects.

4.2 Logistic specification

4.2.1 Matriculation examination score as a dependent variable

The logistic transformation of the dependent variable did not change the basic nature of results as compared to earlier linear specification. This can be seen by comparing results in table 4 to those in table 2. Parents' education and share of female students have systematically a positive effect on performance. Pupils in urban areas have lower scores than students in less urban areas. When using non-pecuniary explanatory variables only, variables controllable by schools or the school administration (teachers' experience and education, class size and school size) were either insignificant or had unexpected signs (teachers' experience) in models explaining matriculation examination scores. As before the picture changed somewhat when either average teachers salaries or teaching expenditures were substituted for non-pecuniary teaching variables. The pecuniary variables got positive and significant coefficients when the dependent variable was the matriculation examination score in 1991 but not in 1992, just as in earlier linear specifications.

The coefficients of linear and logistic specifications are of not comparable as such although their signs are comparable indicating direction of effect. In order to be able to compare the predictions of the two model types we calculated using formulas (6) and (7) the predicted values for matriculation examination score for each school using the untransformed predicted values from the weighted regressions. Then we could calculate the residuals and the R-squares which can be compared pairwise with R-square of the respective linear model. This comparison is not a test of two model specifications, only a comparison of their ability to predict using R-square type indicator. The results indicate that the overall predictive power of the logistic specification is invariably somewhat worse than that of the respective linear models.

Table 4. Estimation results of weighted regression models with logistic specification using the average scores in matriculation examination in 1991 and in 1992 as the dependent variables.

	Average score in matriculation examination 1991					**Average score in matriculation examination 1992**				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**	**Model 1**	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.042458 (1.583)	0.038568 (1.428)	0.062245 (1.617)	0.035991 (1.336)	0.063704* (1.664)	0.055547** (2.034)	0.054351** (1.979)	0.12339*** (3.121)	0.050596* (1.837)	0.1262*** (3.182)
PARENTS' EDUCATION	0.005327*** (13.92)	0.00537*** (14.09)	0.00477*** (9.153)	0.00529*** (13.91)	0.00479*** (9.29)	0.00499*** (14.15)	0.0051*** (14.54)	0.00422*** (7.271)	0.00502*** (14.3)	0.00415*** (7.199)
PERCENTAGE FEMALE	0.82277*** (4.556)	0.85277*** (4.728)	0.532820** (2.04)	0.83121*** (4.607)	0.49100* (1.884)	0.6566*** (3.694)	0.66266*** (3.716)	0.55977** (2.355)	0.66939*** (3.762)	0.55257** (2.319)
TEACHERS' EXPERIENCE	0.011230 (0.3446)					0.026997 (0.7694)				
TEACHERS' EDUCATION	-0.337460** (-2.247)					-0.35021** (-2.174)				
AVERAGE TEACHERS' SALARIES		0.1* (1.826)	0.173460*** (2.69)				0.078313 (1.357)	0.11460* (1.884)		
CLASS SIZE	-0.004324 (-0.8239)	-0.009135* (-1.737)	-0.004854 (-0.7479)			-0.004531 (-0.809)		-0.002717 (-0.4257)		
TEACHING EXPENDI-TURES/PUPIL				0.11085** (2.247)	0.14821*** (2.788)		-0.008581 (-1.572)		0.060936 (1.244)	0.057685 (1.187)
SCHOOL SIZE	-0.079 (-0.4975)	-0.004 (-0.023)	-0.191260 (-0.8316)	0.106010 (0.6382)	0.110910 (0.4474)	0.000001 (0.008)	0.048900 (0.2905)	-0.201630 (-0.8215)	0.102250 (0.5878)	-0.067999 (-0.2664)
URBAN AREA	-0.15082*** (-3.089)	-0.17067*** (-3.478)	-0.13142* (-1.718)	-0.15456*** (-3.2)	-0.13119* (-1.723)	-0.15002*** (-2.864)	-0.16304*** (-3.088)	-0.085638 (-1.117)	-0.15332*** (-2.921)	-0.076192 (-0.9941)
DENSELY POPULATED AREA	-0.030605 (-0.6286)	-0.034709 (-0.7129)	-0.062557 (-1.174)	-0.012437 (-0.252)	-0.028922 (-0.5416)	0.005776 (0.106)	0.005995 (0.1099)	0.014013 (0.2489)	0.013392 (0.2402)	0.027652 (0.4837)
CONSTANT	-1.3503*** (-5.099)	-1.7234*** (-7.162)	-1.7173*** (-5.164)	-1.8654*** (-8.238)	-1.7779*** (-5.477)	-1.3426*** (-5.075)	-1.6534*** (-6.685)	-1.9586*** (-6.144)	-1.7427*** (-7.651)	-1.9078*** (-6.144)
$\tilde{R}^2$	0.173	0.175	0.068	0.202	0.110	0.278	0.279	0.175	0.274	0.162
N	384	384	239	384	239	350	350	205	350	205

***$p < 0.01$ **$p < 0.05$ *$p < 0.10$

4.2.2 Final report score as a dependent variable

Results from employing the logistic specification in models with average final report scores as the dependent variable are given in Table 5. Again the overall picture remains much the same as before. Relative to having matriculation examination score as dependent variable, now also admission level becomes invariably a positive and significant factor. School size now most often has a positive and significant effect on performance. Non-pecuniary measures of school resources are invariably insignificant in all the models with school report scores as dependent variable. The same is true for average teachers' salaries, whereas teaching expenditures per pupil get positive and significant coefficients in 1991 but not in 1992. The R-squares calculated from the results of the logistic specifications are again somewhat lower than those of respective linear specifications.

Table 5. Estimation results of weighted regression models with logistic specification using average final report scores in 1991 and 1992 as dependent variables.

	Average final report score in 1991					Average final report score in 1992				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**	**Model 1**	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.049297*** (3.54)	0.047341*** (3.39)	0.059060*** (2.923)	0.045857*** (3.299)	0.057991*** (2.892)	0.062604*** (4.525)	0.062037*** (4.476)	0.055299*** (2.815)	0.059980*** (4.329)	0.055910*** (2.847)
PARENTS' EDUCATION	0.001751*** (8.814)	0.001758*** (8.923)	0.001895*** (6.927)	0.001726*** (8.798)	0.001871*** (6.932)	0.001368*** (7.661)	0.001389*** (7.849)	0.002186*** (7.583)	0.001362*** (7.713)	0.002159*** (7.558)
PERCENTAGE FEMALE	0.49040*** (5.231)	0.49635*** (5.323)	0.45056*** (3.287)	0.48508*** (5.213)	0.43577*** (3.192)	0.40093*** (4.453)	0.40228*** (4.47)	0.54310*** (4.599)	0.40424*** (4.516)	0.54003*** (4.576)
TEACHERS' EXPERIENCE	0.007173 (0.4239)					0.001037 (0.058)				
TEACHERS' EDUCATION	-0.063034 (-0.8083)					-0.063371 (-0.7767)				
AVERAGE TEACHERS' SALARIES		0.042191 (1.429)	0.049422 (1.46)				0.017861 (0.6132)	0.044860 (1.484)		
CLASS SIZE	-0.001603 (-0.5881)	-0.002874 (-1.057)	-0.004021 (-1.18)			-0.001408 (-0.4961)	-0.002365 (-0.8584)	-0.000981 (-0.3095)		
TEACHING EXPENDI-TURES/PUPIL				0.052634** (2.068)	0.056646** (2.035)				0.272630 (1.107)	0.027420 (1.139)
SCHOOL SIZE	0.152550* (1.854)	0.167040** (2.074)	0.085619 (0.7092)	0.223220** (2.605)	0.182410 (1.405)	0.20756** (2.412)	0.21604** (2.543)	-0.060474 (-0.496)	0.2397*** (2.739)	-0.000225 (-0.002)
URBAN AREA	-0.17126*** (-6.757)	-0.17693*** (-6.975)	-0.13716*** (-3.417)	-0.16826*** (-6.754)	-0.13952*** (-3.497)	-0.17238*** (-6.495)	-0.17569*** (-6.594)	-0.15581*** (-4.091)	-0.17184*** (-6.509)	-0.15189*** (-4.002)
DENSELY POPULATED AREA	-0.085576*** (-3.385)	-0.086965*** (-3.455)	-0.070528** (-2.521)	-0.074678*** (-2.934)	-0.061472** (-2.198)	-0.07744*** (-2.806)	-0.077854*** (-2.829)	-0.058966** (-2.108)	-0.07283*** (-2.597)	-0.052238* (-1.845)
CONSTANT	-0.44019*** (-3.202)	-0.52237*** (-4.199)	-0.59095*** (-3.385)	-0.57613*** (-4.933)	-0.66785*** (-3.928)	-0.3796*** (-2.833)	-0.44976*** (-3.603)	-0.76291*** (-4.817)	-0.49448*** (-4.316)	-0.74808**** (-4.865)
$\tilde{R}^2$	0.188	0.186	0.106	0.210	0.135	0.224	0.223	0.180	0.222	0.177
N	384	384	239	384	239	350	350	205	350	205

***$p < 0.01$ **$p < 0.05$ * $p < 0.10$

4.3 Results of models explaining variability of performance

In sections 4.1 and 4.2 we have reported results from models in which the dependent variable was an average performance indicator of each school, either matriculation examination score or final report score. Thus no attention was paid to the eventual effect of independent variables on variability of performance. Although at the ends of performance distribution there is no variability (all fail or all are excellent), in the middle it is possible that some factor may affect either similarly or the opposite on variation of performance as on average performance. Furthermore, there may be no effect on average performance, only a change in variability. To test this kind of effects we, following Brown and Saks (1975), used the standard deviation of our two performance indicators as the dependent variables and ran linear weighted OLS regressions with the same independent variables, data alternatives and model types as before. The results of these regressions are in tables 6 and 7.

Referring to table 6, only a few variables get statistically significant coefficients when the standard deviation of matriculation examination score is used as the dependent variable. Parents' education which affected average performance positively in all previous models now has a negative effect on standard deviation of performance in several but not all models. With the exception of admission level and teachers' average salary which get a negative an significant coefficients in one out of ten variants all other variables are invariable insignificant.

The respective results with standard deviation of final report score as the dependent variable are different. Now parents' education is never significant at 5 per cent level. Instead admission level gets invariably a negative and significant coefficient. Furthermore, school size gets a positive and one of the area dummies (densely populated area) in some models a negative and significant coefficient.

To conclude the independent variables have even smaller explanatory power for the variability of performance. The R-squares were low, at the highest 0,12 when the standard deviation of matriculation examination was used as the dependent variable. When the standard deviation of final report scores was used as the dependent variable the R-squares were at best 0,04. In other words only 4 per cent of the variability could be explained.

Table 6. Estimation results of weighted linear regression models using the standard deviation of matriculation examination scores in 1991 and 1992 as dependent variables.

	Standard deviation of matriculation examination scores in 1991					**Standard deviation of matriculation examination scores in 1992**				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**	**Model 1**	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	-0.0060004 (-0.3694)	-0.0069960 (-0.4252)	-0.022312 (-0.9809)	-0.0030097 (-0.1829)	-0.021627 (-0.9572)	-0.027942* (-1.742)	-0.026494* (-1.645)	-0.048517** (-2.012)	-0.025602 (-1.589)	-0.050973** (-2.102)
PARENTS' EDUCATION	-0.001415*** (-6.104)	-0.00144*** (-6.201)	-0.00155*** (-5.039)	-0.001386*** (-5.968)	-0.00154*** (-5.052)	-0.00143*** (-6.905)	-0.00148*** (-7.2)	-0.00120*** (-3.397)	-0.00146*** (-7.086)	-0.00116*** (-3.274)
PERCENTAGE FEMALE	-0.19880* (-1.818)	-0.21490** (-1.956)	-0.230990 (-1.496)	-0.21168* (-1.922)	-0.215740 (-1.402)	-0.25494** (-2.442)	-0.2576** (-2.463)	-0.29261** (-2.018)	-0.25326** (-2.433)	-0.28768** (-1.974)
TEACHERS' EXPERIENCE	0.011538 (0.5846)					0.002768 (0.1343)				
TEACHERS' EDUCATION	0.19142** (2.104)					0.155120 (1.64)				
AVERAGE TEACHERS' SALARIES		0.0039977 (0.1149)	-0.053818 (-1.412)				-0.039963 (-1.181)	-0.075405** (-2.032)		
CLASS SIZE	0.0035937 (1.131)	0.0054197* (1.692)	0.003417 (0.8908)			-0.001045 (-0.3176)	0.001430 (0.4469)	0.002366 (0.6078)		
TEACHING EXPENDI-TURES/PUPIL				-0.036105 (-1.199)	-0.057221* (-1.824)				-0.032795 (-1.145)	-0.030434 (-1.024)
SCHOOL SIZE	0.11182 (1.165)	0.069538 (0.7329)	0.040635 (0.2989)	0.041239 (0.4067)	-0.062388 (-0.4264)	-0.033513 (-0.3359)	-0.055289 (-0.5601)	0.016198 (0.1082)	-0.089692 (-0.8815)	-0.057399 (-0.3678)
TOWNLIKE AREA	-0.0062342 (-0.2109)	0.00017189 (0.005751)	0.026828 (0.5935)	0.0044884 (0.1522)	0.028572 (0.6356)	0.016197 (0.5263)	0.024519 (0.7919)	-0.041766 (-0.8931)	0.015147 (0.4934)	-0.047283 (-1.009)
DENSELY POPULATED AREA	-0.020333 (-0.6896)	-0.019288 (-0.6504)	0.020696 (0.6571)	-0.018966 (-0.6296)	0.010505 (0.3333)	-0.025248 (-0.789)	-0.023809 (-0.7443)	-0.019702 (-0.5737)	-0.033588 (-1.03)	-0.025863 (-0.7399)
CONSTANT	1.5975*** (9.961)	1.7825*** (12.16)	2.0829*** (10.6)	1.9229*** (13.91)	2.1461*** (11.2)	1.9683*** (12.67)	2.1473*** (14.8)	2.284*** (11.75)	2.147*** (16.12)	2.2481*** (11.84)
$\tilde{R}^2$	0.057	0.054	0.045	0.061	0.067	0.117	0.114	0.098	0.108	0.083
N	384	384	239	384	239	350	350	205	350	205

***$p < 0.01$ **$p < 0.05$ * $p < 0.10$

Table 7. Estimation results of weighted linear regression models using the standard deviation of final report scores in 1991 and 1992 as dependent variables.

	Standard deviation of average final report scores in 1991					Average final report score in 1992				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**	**Model 1**	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	-0.02059*** (-2.249)	-0.02147** (-2.317)	-0.0289** (-2.268)	-0.02272** (-2.461)	-0.03091** (-2.431)	-0.0398*** (-4.263)	-0.03982*** (-4.255)	-0.01978 (-1.441)	-0.04002*** (-4.275)	-0.020924 (-1.524)
PARENTS' EDUCATION	-0.000200 (-1.53)	-0.00022* (-1.6979)	-0.000218 (-1.261)	-0.00023* (-1.761)	-0.000235 (-1.374)	0.000085 (0.7056)	0.000070 (0.5861)	-0.000110 (-0.546)	0.000072 (0.5998)	-0.000106 (-0.532)
PERCENTAGE FEMALE	-0.075878 (-1.231)	-0.085529 (-1.383)	-0.115080 (-1.331)	-0.089318 (-1.445)	-0.116150 (-1.342)	-0.089488 (-1.473)	-0.089598 (-1.474)	-0.1568* (-1.900)	-0.089657 (-1.483)	-0.15649* (-1.897)
TEACHERS' EXPERIENCE	-0.003393 (-0.305)					-0.001817 (-0.1515)				
TEACHERS' EDUCATION	0.11554** (2.254)					0.050019 (0.9085)				
AVERAGE TEACHERS' SALARIES		-0.002343 (-0.1196)	-0.001599 -0.7489)				-0.005136 (-0.2611)	-0.023858 (-1.130)		
CLASS SIZE	-0.000459 (-0.2561)	0.000391 (0.2169)	-0.001031 (-0.4798)			-0.000204 (-0.1066)		-0.000190 (-0.086)		
TEACHING EXPENDI-TURES/PUPIL				0.015240 (0.9017)	0.007831 (0.4436)		0.000333 (0.1792)		0.000563 (0.033860)	-0.009161 (-0.544)
SCHOOL SIZE	-0.073890 (-1.366)	-0.09768* (-1.829)	-0.12650* (-1.661)	-0.077984 (-1.371)	-0.122530 (-1.488)	-0.11897** (-2.049)	-0.12680** (-2.211)	-0.096054 (-1.127)	-0.12655** (-2.141)	-0.122760 (-1.389)
TOWNLIKE AREA	0.03141* (1.885)	0.03425** (2.036)	0.04425* (1.748)	0.03629** (2.194)	0.04191* (1.657)	0.018269 (1.02)	0.019659 (1.093)	0.000534 (0.020)	0.019451 (1.091)	-0.001946 (-0.073)
DENSELY POPULATED AREA	0.017079 (1.028)	0.016740 (1.003)	0.030448 (1.726)	0.020939 (1.239)	0.02939* (1.657)	-0.006327 (-0.3398)	-0.006298 (-0.3389)	0.004261 (0.218)	-0.005812 (-0.3067)	0.001071 (0.054)
CONSTANT	1.0665*** (11.8)	1.1612*** (14.07)	1.2833*** (11.65)	1.1451*** (14.76)	1.2354*** (11.46)	1.1907*** (13.17)	1.2316*** (14.61)	1.2323*** (11.14)	1.2277*** (15.86)	1.209*** (11.25)
$\tilde{R}^2$	0.010	0.015	0.024	0.016	0.018	0.040	0.041	0.033	0.039	0.030
N	384	384	239	384	239	350	350	205	350	205

$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.10$

4.4 The results using pooled data

To conclude, in addition to the above data alternatives, we also estimated the equations with a pooled data containing both years. Here each school was twice as an observation such that the dependent variable and most of the independent variables had different values for the two years. Only the area dummies were the same for both years. The models included also a dummy variable indicating the year (1991=0 and 1992=1).

The results of pooled data using linear specification and average performance measures are reported in Appendix 3. Qualitatively the main results remain much the same as before. The R-squares indicating the overall explanatory power of the models as well as the t-values indicating statistical significance of each explanatory variables were higher than in the data constructed on yearly basis. The most significant difference to the yearly data was that in the case of matriculation examination scores in model 2 and model 3 also schools' expenditures per pupil had positive and statistically significant coefficients in addition to students' background, schools' admission level and the area, especially when DATA 1 was used. This effect was more obscure in the case of final report scores where the year-dummy in turn had a negative and statistically significant coefficient. It means that when students' background, the resources and the area of school are controlled, students who graduated in 1992 had lower final report scores than those who graduated in 1991.

The estimation results of logistic specification with pooled data are reported in the Appendix 4. The results did not change compared to estimations based on yearly data. The only difference was also in this case that the t-values were somewhat stronger than in the estimations using yearly data. Also in this case the year-dummy got a negative statistically significant coefficient when the final report score was used as the dependent variable.

Again the use of pooled data in the case of explaining the variability of performance did only change the t-values of coefficients so that they were stronger. As opposed to explaining the average performance the year-dummy did not get a statistically significant sign when the standard deviation of final report score was used as a dependent variable. The results of estimations using pooled data are in the Appendix 5.

5 CONCLUSIONS

In this paper we have studied how the resources of senior secondary schools affect the average achievement level of their pupils. In our regression analyses the dependent variables were a score based on pupils' average performance in national matriculation examinations and a average score in final reports given by the teachers in the school. Separate regressions were run for two years, namely average scores of 1991 and 1992 for both of the variables. The independent variables included both input variables controllable by the schools or the school administration, and variables related to the socioeconomic background of the pupils.

On the basis of earlier research the relation between school inputs and outcomes is somewhat ambiguous. Rather often school related variables do not have statistical power in explaining variability in achievement levels. The socioeconomic background of pupils, on the contrary, seems to be an important factor behind achievement level although the variables used are typically only proxies for income or education related factors which are important for good performance. This kind of results would mean that the possibilities to affect achievement by such factors that are controllable by the schools or school administration are rather limited.

The results of this study are similar to those of earlier international evidence. In all the models the quality of students measured by admission level (measured by lowest mark among accepted pupils at the time of entrance) and their socioeconomic background (measured by the average education level of students' parents) turned out to have a positive effect on achievement. The share of female students to had a positive effect on school performance in most of the models, too. Furthermore, in cities the achievement levels are lower than in less urban areas. Surprisingly, class size or teachers' experience and education either had no effect or it had an unexpected effect. School size sometimes had a positive effect on achievement when the final report scores were used as output measure.

When the above mentioned school resources were measured in terms of money the results were somewhat different. Teachers' average salaries and other teaching expenditures per pupil had a statistically significant positive impact on average achievement level in some models.

Above, we have discussed results from models in which the dependent variable was an average performance indicator. Thus no attention was paid to the eventual effect of independent variables on variability of performance. To test this kind of effects we used the standard deviation of our two performance measures as the dependent variables and ran linear OLS regressions with the same independent variables, data alternatives and model types as before. According to our results, parents education which affected average performance positively in all previous models now has a negative effect on standard deviation of performance in several but not all models. With the exception of admission level and teachers' average salary which get a negative an significant coefficient in one out of tens variants, all other variables are invariable insignificant.

The respective results with final report scores are different. Now parents' education is never significant, instead admission level gets invariably a negative and significant coefficient. Furthermore, school size gets a positive and one of the area variables (densely populated area) a negative and significant coefficient in some models. In other words the resource variables controlled by schools and school administration did neither affect the variability of performance.

Like in earlier studies on this topic, most of the variation in achievement levels of schools remained statistically unexplained. The R-squares were rather low, at best around thirty per cent. This is an indication that there are a number of factors that are either absent from our models or the variables used are not properly operationalized. In other words, there is much room for further work. Nevertheless, if the most often heard assertions that with better resources and smaller class sizes one gets better school results were unquestionably true, one would have expected them to emerge already in our analysis.

REFERENCES

Alexander, L. and J. Simmons. The Determinants of School Achievement in Developing Countries: The Educational Production Function. *Staff Working Paper* no. 201. Washington DC: IBRD, 1975.

Bonesronning, H. and J. Rattso. "Efficiency Variation Among the Norwegian High Schools: Consequences of Equalization Policy," *Economics of Education Review* 13 (1994): 289-304.

Brown, B and D. Saks. "The Production and Distribution of Cognitive Skills within Schools," *Journal of Political Economy* 83 (1975): 571-593.

Card, D. and A. B. Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100 (1992): 1-40.

Cohn, E. and T.G. Geske. *Economics of Education*. Oxford: Pergamon Press, 1990.

Cuttance, P. Frameworks for Research on the Effectiveness of Schooling. In D. Reynolds (ed.) *Studying School Effectiveness*. East Sussex: The Falmer Press, 1985.

Dolan, R. C and R. M. Schmidt. "Assessing the Impact of Expenditure on Achievement: Some Methodological and Policy Considerations," *Economics of Education Review* 6 (1987): 285-299.

Ehrenberg, R. G. and D. J. Brewer. "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond," *Economics of Education Review* 13 (1994): 1-17.

Finansdepartementet. Kostnader och resultat i Grundskolan, - en Jämförelse av Kommuner. *Rapport till Expertgruppen för Studier i Offentlig Ekonomi*, Ds Fi 14, 1986.

Fuller, B. What School Factors Raise Achievement in the Third World? *Review of Educational Research* 57 (1987): 255-292.

Greene, W. H. *Econometric Analysis*. Second edition. New York: Macmillan Publishing Company, 1993.

Hanushek, E. A. "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 26 (1986): 1141-1177.

Hanushek, E. A. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *The Journal of Human Resources* 14 (1979): 351-388.

Hanushek, E. A. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *American Economic Review* 61 (1971): 280-288.

Kirjavainen, T. and H. A. Loikkanen. On the Efficiency Differences of Finnish Senior Secondary Schools-An Application of DEA for Measuring the Efficiency Differences (in Finnish), *VATT-research reports* 16. Helsinki, 1993.

Levine D. U. An Interpretive Review of US Research and Practice Dealing with Unusually Effective Schools. In D. Reynolds and P. Cuttance (eds.) *School Effectiveness. Research, policy and practice*. London: Cassell, 1992.

Lewis, K. A. and L. S. Seidman. "Math-Time Capital Matters: A Cross-Country Analysis," *Economics of Education Review* 13 (1994): 215-226.

Plowden Report. Children and Their Primary Schools. Report of the Central Advisory Council on Education. London: Her Majesty's Stationery Office, 1967.

Purkey, S. C. and M. S. Smith. "Effective Schools: A Review," *The Elementary School Journal* 83 (1983): 427-452.

Reynolds, D. School Effectiveness and School Improvement: An Updated Review of the British Literature. In D. Reynolds and P. Cuttance (eds.) *School Effectiveness. Research, policy and practice*. London: Cassell, 1992.

Sandberg, S. Effektiv grundskola, Servicetillfredställelse, serviceproduction och produktivitet i 36 kommuner. Turku: Åbo Akademi, 1992.

Sander, W. "Expenditures and Student Achievement in Illinois," *Journal of Public Economics* 52 (1993): 403-416.

Simmons, J. How Effective is Schooling in Promoting Learning? Review of the Literature. *Staff Working Paper no. 200*. Washington DC: IBRD, 1975.

APPENDIX 1. The formula used for correcting teachers' salaries from regional differences.

Teachers' salaries were corrected from the regional differences by using the following formula:

I group

salaries=salaries/(1.028(1+0.04*number of points for remoteness))

II group
salaries=salaries /(1+0.04* number of points for remoteness)

APPENDIX 2. Summary statistics of the variables.

DATA 1 used for the year 1991 estimations.

NAME	N	MEAN	ST. DEV.	MINIMUM	MAXIMUM
MATRICULATION EXAM. SCORE	384	4.2289	0.39881	2.944	5.604
FINAL REPORT SCORE	384	7.8598	0.24811	7.01	8.94
STD. DEV. OF MATRICULATION EXAM. SCORE	384	1.2945	0.20536	0.63960	1.8157
STD. DEV. OF FINAL REPORT SCORE	384	0.90161	0.11384	0.45000	1.3000
ADMISSION LEVEL	384	6.8578	0.58917	5	9.09
PARENTS' EDUCATION	384	301.32	51.405	0	531.6
PERCENTAGE FEMALE	384	0.58927	9.87E-02	0.29412	1
TEACHERS' EDUCATION	384	0.84004	0.11782	0.283	1
TEACHERS' EXPERIENCE	384	3.4536	0.6041	1.222	4.804
AVERAGE TEACHERS' SALARIES	384	207000	29236	66914	423000
CLASS SIZE	384	19.981	4.1208	5.592	58.929
TEACHING EXPENDITURES /PUPIL	384	18414	5174.3	11516	61187
SCHOOL SIZE	384	182.03	100.1	23	622
DENSELY POPULATED AREA	384	0.1849	0.38872	0	1
URBAN AREA	384	0.42708	0.4953	0	1

APPENDIX 2. Summary statistics of the variables.

Data 1 used for the 1992 estimations.

NAME	N	MEAN	ST. DEV.	MINIMUM	MAXIMUM
MATRICULATION EXAM. SCORE	350	4.2618	0.41198	3.105	5.476
FINAL REPORT SCORE	350	7.8567	0.23695	7.18	8.89
STD. DEV. OF MATRICULATION EXAM. SCORE	350	1.2874	0.20900	0.59630	1.8873
STD. DEV. OF FINAL REPRORT SCORE	350	0.91071	0.11456	0.49000	1.2400
ADMISSION LEVEL	350	6.8709	0.59813	5	8.91
PARENTS' EDUCATION	350	317.06	52.775	210.29	540.05
PERCENTAGE FEMALE	350	0.59019	0.10036	0.3	0.88889
TEACHERS EDUCATION	350	0.83634	0.11673	0.2738	1
TEACHERS EXPERIENCE	350	3.4432	0.57686	1.333	4.651
AVERAGE TEACHER SALARIES	350	206000	29769	62057	453000
CLASS SIZE	350	20.148	4.0438	5.759	58.537
TEACHING EXPENDITURES /PUPIL	350	18158	5874.3	11581	78535
SCHOOL SIZE	350	190.66	100.89	24	623
DENSELY POPULATED AREA	350	0.17143	0.37742	0	1
URBAN AREA	350	0.47143	0.4999	0	1

APPENDIX 2. Summary statistics of the variables.

Data 2 used for the 1991 estimations.

NAME	N	MEAN	ST. DEV.	MINIMUM	MAXIMUM
MATRICULATION EXAM. SCORE	239	4.1944	0.3897	3.08	5.604
FINAL REPORT SCORE	239	7.8939	0.25082	7.08	8.94
STD. DEV. OF MATRICULATION EXAM. SCORE	239	1.3145	0.21198	0.63960	1.8157
STD. DEV. OF FINAL REPRORT SCORE	239	0.90038	0.12019	0.4500	1.3000
ADMISSION LEVEL	239	6.7916	0.54045	5	8.27
PARENTS' EDUCATION	239	285.57	50.193	0	531.6
PERCENTAGE FEMALE	239	0.60092	0.0948	0.29412	1
TEACHERS EDUCATION	239	0.83379	0.12426	0.283	1
TEACHERS EXPERIENCE	239	3.3716	0.66281	1.222	4.804
AVERAGE TEACHER SALARIES	239	201000	32067	66914	423000
CLASS SIZE	239	19.328	4.6415	5.592	58.929
TEACHING EXPENDITURES /PUPIL	239	19320	6211.1	11516	61187
SCHOOL SIZE	239	154.79	91.355	23	622
DENSELY POPULATED AREA	239	0.22176	0.4163	0	1
URBAN AREA	239	0.17155	0.37778	0	1

APPENDIX 2. Summary statistics of the variables.

Data 2 for the 1992 estimations.

NAME	N	MEAN	ST. DEV.	MINIMUM	MAXIMUM
MATRICULATION EXAM. SCORE	205	4.2195	0.4003	3.105	5.476
FINAL REPORT SCORE	205	7.8881	0.23508	7.25	8.89
STD. DEV. OF MATRICULATION EXAM. SCORE	205	1.3064	0.22072	0.59360	1.8873
STD. DEV. OF FINAL REPRORT SCORE	205	0.90873	0.11919	0.4900	1.2400
ADMISSION LEVEL	205	6.8224	0.57333	5	8.64
PARENTS' EDUCATION	205	297.9	43.062	210.29	540.05
PERCENTAGE FEMALE	205	0.60306	0.0990	0.35484	0.88889
TEACHERS' EDUCATION	205	0.83724	0.11888	0.2738	1
TEACHERS' EXPERIENCE	205	3.3736	0.6333	1.333	4.651
AVERAGE TEACHER' SALARIES	205	201000	34230	62057	453000
CLASS SIZE	205	19.515	4.6889	5.759	58.537
TEACHING EXPENDITURES /PUPIL	205	19190	7393	11581	78535
SCHOOL SIZE	205	157.14	90.028	24	623
DENSELY POPULATED AREA	205	0.22439	0.4182	0	1
URBAN AREA	205	0.17561	0.38142	0	1

APPENDIX 3. The estimation results of weighted linear regression models using pooled data and average matriculation examination score as the dependent variable.

	Average matriculation examination score				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.052401*** (2.549)	0.049772** (2.405)	0.091894*** (3.090)	0.046622** (2.248)	0.093804*** (3.148)
PARENTS' EDUCATION	0.00497*** (17.90)	0.005056*** (18.27)	0.0042*** (10.07)	0.004973*** (17.94)	0.004164*** (10.04)
PERCENTAGE FEMALE	0.75464*** (5.562)	0.77484*** (5.696)	0.51825*** (2.722)	0.77009*** (5.658)	0.49816*** (2.609)
TEACHERS' EXPERIENCE	0.015795 (0.6155)				
TEACHERS' EDUCATION	-0.36590*** (-3.101)				
AVERAGE TEACHERS' SALARIES		0.94001** (2.156)	0.14524*** (3.033)		
CLASS SIZE	-0.0047645 (-1.158)	-0.0095342** (-2.342)	-0.0041277 (-0.8384)		
TEACHING EXPENDI-TURES/PUPIL				0.83709** (2.239)	0.95148** (2.442)
SCHOOL SIZE	-0.0045626 (-0.03662)	0.061094 (0.4967)	-0.16434 (-0.9093)	0.013619 (1.054)	0.032775 (0.1709)
URBAN AREA	-0.13845*** (-3.611)	-0.15642*** (-4.049)	-0.10455* (-1.783)	-0.14379*** (-3.760)	-0.097372* (-1.661)
DENSELY POPULATED AREA	-0.0010810 (-0.02770)	-0.0036039 (-0.09224)	-0.010879 (-0.2601)	0.010295 (0.2584)	0.011916 (0.2816)
YEAR	-0.036844 (-1.556)	-0.034090 (-1.437)	-0.011607 (-0.3652)	-0.034362 (-1.447)	-0.012434 (-0.3904)
CONSTANT	2.3289*** (11.60)	1.9533*** (10.55)	1.8781*** (7.561)	1.8347*** (10.63)	1.8887*** (7.783)
$\tilde{R}^2$	0.251	0.251	0.153	0.256	0.155
N	734	734	444	734	444

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

APPENDIX 3. The estimation results of weighted linear regression models using pooled data and average final report score as the dependent variable.

	Average final report score				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.07386*** (5.635)	0.072156*** (5.490)	0.07531*** (4.090)	0.06996*** (5.327)	0.074636*** (4.060)
PARENTS' EDUCATION	0.0019737*** (11.15)	0.0019937*** (11.35)	0.0024686** (9.552)	0.0019533** (11.13)	0.0024338** (9.509)
PERCENTAGE FEMALE	0.58476*** (6.758)	0.59101*** (6.841)	0.65290*** (5.537)	0.58582*** (6.798)	0.64190*** (5.449)
TEACHERS' EXPERIENCE	0.0039309 (0.2402)				
TEACHERS' EDUCATION	-0.088302 (-1.174)				
AVERAGE TEACHERS' SALARIES		0.41472 (1.498)	0.59674* (2.012)		
CLASS SIZE	-0.0019050 (-0.7262)	-0.0034674 (-1.341)	-0.0032411 (-1.063)		
TEACHING EXPENDI-TURES/PUPIL				0.50399* (2.129)	0.51063* (2.125)
SCHOOL SIZE	0.26024*** (3.275)	0.27543*** (3.526)	0.054208 (0.4843)	0.32484*** (3.971)	0.14775 (1.248)
URBAN AREA	-0.22789*** (-9.320)	-0.23453*** (-9.560)	-0.19365*** (-5.333)	-0.22599*** (-9.333)	-0.19198*** (-5.308)
DENSELY POPULATED AREA	-0.10622*** (-4.269)	-0.10757*** (-4.335)	-0.08267*** (-3.191)	-0.09644*** (-3.824)	-0.07273*** (-2.786)
YEAR	-0.0345** (-2.285)	-0.033479** (-2.221)	-0.043594** (-2.215)	-0.033541** (-2.230)	-0.043586** (-2.218)
CONSTANT	6.5816*** (51.41)	6.4686*** (55.00)	6.2709*** (40.76)	6.4108*** (58.69)	6.2381*** (41.67)
$\tilde{R}^2$	0.213	0.211	0.152	0.220	0.164
N	734	734	444	734	444

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

APPENDIX 4. The estimation results of logistic regressions with pooled data using the number of students taking matriculation examination as weights.

	Average matriculation examination score				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.049327*** (2.595)	0.046911** (2.452)	0.09158*** (3.357)	0.043709** (2.281)	0.09306*** (3.405)
PARENTS' EDUCATION	0.00513*** (19.99)	0.00521*** (20.38)	0.00454*** (11.86)	0.00513*** (20.05)	0.00451*** (11.86)
PERCENTAGE FEMALE	0.73593*** (5.865)	0.75512*** (6.004)	0.55106*** (3.154)	0.74904*** (5.955)	0.53077*** (3.031)
TEACHERS' EXPERIENCE	0.019144 (0.8067)				
TEACHERS' EDUCATION	-0.34327*** (-3.146)				
AVERAGE TEACHERS' SALARIES		0.09130** (2.265)	0.01422*** (3.235)		
CLASS SIZE	-0.004342 (-1.141)	-0.00872** (-2.318)	-0.003535 (-0.7824)		
TEACHING EXPENDI-TURES/PUPIL				0.08522** (2.467)	0.0989*** (2.765)
SCHOOL SIZE	-0.039953 (-0.3467)	0.020878 (0.1836)	-0.208090 (-1.255)	0.099329 (0.8319)	-0.002299 (-0.013)
URBAN AREA	-0.14974*** (-4.223)	-0.16635*** (-4.658)	-0.10900** (-2.026)	-0.15275*** (-4.323)	-0.10178* (-1.893)
DENSELY POPULATED AREA	-0.014051 (-0.3894)	-0.016058 (-0.4446)	-0.024422 (-0.6363)	-0.000548 (-0.015)	-0.000238 (-0.006)
YEAR	-0.03963* (-1.81)	-0.03720* (-1.696)	-0.021149 (-0.7253)	-0.03748* (-1.708)	-0.021931 (-0.7507)
CONSTANT	-1.3218*** (-7.12)	-1.6682*** (-9.743)	-1.8353*** (-8.052)	-1.7797*** (-11.16)	-1.8261*** (-8.204)
$\tilde{R}^2$	0.226	0.226	0.118	0.234	0.122
N	734	734	444	734	444

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

APPENDIX 4. Estimation results of logistic regressions with pooled data using the number of students taking matriculation examination as weights.

	Average final report score				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	0.05641*** (5.778)	0.05525*** (5.644)	0.05886*** (4.208)	0.05344*** (5.467)	0.05819*** (4.171)
PARENTS' EDUCATION	0.00154*** (11.66)	0.00155*** (11.86)	0.00202*** (10.28)	0.00152*** (11.64)	0.00199*** (10.25)
PERCENTAGE FEMALE	0.44054*** (6.836)	0.44505*** (6.917)	0.50219*** (5.608)	0.44089*** (6.872)	0.49366*** (5.521)
TEACHERS' EXPERIENCE	0.004925 (0.4041)				
TEACHERS' EDUCATION	-0.062666 (-1.118)				
AVERAGE TEACHERS' SALARIES		0.030552 (1.481)	0.044782** (1.988)		
CLASS SIZE	-0.001494 (-0.7647)	-0.002548 (-1.324)	-0.002370 (-1.024)		
TEACHING EXPENDI-TURES/PUPIL				0.039642** (2.25)	0.040235** (2.206)
SCHOOL SIZE	0.178020*** (3.008)	0.188440*** (3.239)	0.026240 (0.3087)	0.22756*** (3.736)	0.100170 (1.115)
URBAN AREA	-0.17027*** (-9.35)	-0.17481*** (-9.568)	-0.14702*** (-5.331)	-0.16815*** (-9.329)	-0.14576*** (-5.31)
DENSELY POPULATED AREA	-0.081*** (-4.37)	-0.08176*** (-4.425)	-0.06585*** (-3.347)	-0.07286*** (-3.881)	-0.05791*** (-2.922)
YEAR	-0.02648** (-2.354)	-0.02584** (-2.302)	-0.03443** (-2.304)	-0.02586** (-2.31)	-0.0344** (-2.307)
CONSTANT	-0.39519*** (-4.145)	-0.47195*** (-5.388)	-0.6665*** (-5.705)	-0.51767*** (-6.366)	-0.69295*** (-6.098)
$\tilde{R}^2$	0.206	0.205	0.132	0.214	0.144
N	734	734	444	734	444

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

APPENDIX 5. Results of linear regressions with pooled data using the standard deviation of performance as dependent variable and number of students taking matriculation examination as weights.

	Standard deviation of matriculation examination score				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	-0.016995 (-1.494)	-0.016469 (-1.435)	-0.034625** (-2.114)	-0.013862 (-1.208)	-0.035228** (-2.148)
PARENTS' EDUCATION	-0.00143*** (-9.321)	-0.00147*** (-9.567)	-0.0014*** (-6.075)	-0.00143*** (-9.34)	-0.00137*** (-5.991)
PERCENTAGE FEMALE	-0.2285*** (-3.044)	-0.23587*** (-3.127)	-0.26488** (-2.525)	-0.23551*** (-3.128)	-0.25526** (-2.429)
TEACHERS' EXPERIENCE	0.007963 (0.5608)				
TEACHERS' EDUCATION	0.16849*** (2.581)				
AVERAGE TEACHERS' SALARIES		-0.017222 (-0.7123)	-0.06536** (-2.477)		
CLASS SIZE	0.001337 (0.5874)	0.003524 (1.561)	0.002939 (1.083)		
TEACHING EXPENDI-TURES/PUPIL				-0.03523* (-1.704)	-0.042736** (-1.993)
SCHOOL SIZE	0.000043 (0.6268)	0.009777 (0.1433)	0.038259 (0.3842)	-0.021062 (-0.2947)	-0.044278 (-0.4194)
URBAN AREA	0.005969 (0.2813)	0.013279 (0.62)	-0.006438 (-0.1993)	0.010730 (0.5072)	-0.008966 (-0.2779)
DENSELY POPULATED AREA	-0.022546 (-1.044)	-0.020726 (-0.9567)	0.000710 (0.031)	-0.025667 (-1.165)	-0.008132 (-0.3491)
YEAR	0.001238 (0.095)	-0.000276 (-0.021)	-0.009061 (-0.5175)	-0.000593 (-0.045)	-0.008807 (-0.5024)
CONSTANT	1.7846*** (16.07)	1.9603*** (19.09)	2.1889*** (15.99)	2.0373*** (21.35)	2.1983*** (16.46)
$\tilde{R}^2$	0.080	0.076	0.063	0.078	0.060
N	734	734	444	734	444

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

APPENDIX 5. Results of linear regressions with pooled data using the standard deviation of performance as dependent variable and number of students taking matriculation examination as weights.

	Standard deviation of final report score				
	Model 1	**Model 2**	**Model 2**	**Model 3**	**Model 3**
	DATA 1	DATA 1	DATA 2	DATA 1	DATA 2
Variable	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)	coefficient (t-value)
ADMISSION LEVEL	-0.03057*** (-4.694)	-0.03093*** (-4.718)	-0.02459*** (-2.664)	-0.03166*** (-4.831)	-0.00263*** (-2.843)
PARENTS' EDUCATION	-0.000041 (-0.4682)	-0.000061 (-0.6946)	-0.000160 (-1.237)	-0.000064 (-0.7319)	-0.000171 (-1.327)
PERCENTAGE FEMALE	-0.081883* (-1.905)	-0.085149** (-1.976)	-0.136540** (-2.311)	-0.086344** (-2.007)	-0.13639** (-2.303)
TEACHERS' EXPERIENCE	-0.002909 (-0.3578)				
TEACHERS' EDUCATION	0.08182** (2.189)				
AVERAGE TEACHERS' SALARIES		-0.003833 (-0.2775)	-0.020147 (-1.356)		
CLASS SIZE	-0.000368 (-0.2823)	0.000341 (0.2647)	-0.000605 (-0.3958)		
TEACHING EXPENDI-TURES / PUPIL				0.007696 (0.6514)	-0.006820 (-0.05645)
SCHOOL SIZE	-0.095335** (-2.415)	-0.11069*** (-2.841)	-0.11288** (-2.012)	-0.10192** (-2.496)	-0.12465** (-2.095)
URBAN AREA	0.024664** (2.03)	0.026653** (2.178)	0.023446 (1.288)	0.027636** (2.287)	0.021399 (1.177)
DENSELY POPULATED AREA	0.005398 (0.4366)	0.005341 (0.4316)	0.018406 (1.418)	0.007754 (0.616)	0.016406 (1.25)
YEAR	0.008405 (1.12)	0.008090 (1.076)	0.006548 (0.6639)	0.008213 (1.094)	0.006978 (0.7063)
CONSTANT	1.1249*** (17.69)	1.1896*** (20.28)	1.2557*** (16.29)	1.1791*** (21.63)	1.2212*** (16.23)
$\tilde{R}^2$	0.028	0.030	0.029	0.025	0.027
N	734	734	444	734	444

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$