MODELLING THE VARIABLES THAT INFLUENCE THE SCHOOL PERFORMANCE IN THE MADRID MUNICIPALITIES

Rocio Gallego
Department of Finance Economy and Accounting II
University Rey Juan Carlos
Paseo Artilleros s/n, 28032 Madrid, Spain
E-mail: rocio.gallego@urjc.es

Miguel Ángel Alonso
Luis Pires
Alberto Romero
Department of Applied Economics I
University Rey Juan Carlos
Paseo Artilleros s/n, 28032 Madrid, Spain
E-mail: miguelangel.alonso@urjc.es; luis.pires@urjc.es; alberto.romero@urjc.es

KEYWORDS
Educational Efficiency, Production Function, School Performance, Madrid Municipalities, Primary School Test Scores

JEL: H52, H75, I21

ABSTRACT
The evaluation of the school performance is a very important strategic field due to the great amount of public and private resources invested by the society in education. State and regional governments undergo several tests and surveys which try to measure the levels of learning attained by their students. While there is a rich literature on the main inputs that explain the school performance at the State level, the concrete determinants at the municipality level—as immigration and the local budget in education—have not been widely addressed.

This paper estimates a cross-section data model using the results of the specific test developed by the Madrid Regional Government in the academic course 2007-08. Our cross-sectional dataset includes 617 schools in 130 municipalities in the surrounding area of Madrid, using data at municipal and school levels. From our study we detect two variables that have an important impact in the scholastic results, the specific characteristics of each school, where the number of foreign students, the ownership of the school, and the municipality expenditure effort in education stand up. The analysis shows the capacity of regional and local government to overcome the education outcomes.

INTRODUCTION
The economic analysis of education has studied the microeconomic effects of education in the labor markets and the workers income, and the macroeconomic influence of education in the economic growth of nations. Nevertheless there is a specific literature that focuses on the analysis of the educational system itself, and the efficiency of schools to educate and bring up students. This literature looks education and schools as a production function, where the inputs, as the students’ personal characteristics and their socio-economic and school environment, produce an output of education achievement. The production function approach to education establishes two main groups of inputs, the personal and environment characteristics of the students (intelligence, family income and culture), and the characteristics of the school and the teachers. Within this second group of characteristics, the government tries to improve the students’ achievement by giving more resources to schools.

In 1966, the U.S. government promoted a study about the educational resources in the country, known as the “Coleman Report” (Coleman et al. 1966). This study, besides describing the school resources in the U.S., used the education production function to measure the influence of different inputs in the students’ achievement. The conclusion of this study was that differences in school resources have little influence in students’ outcome, whilst family background and peers’ characteristics are more important. The “Coleman report” was the beginning of a literature that uses the production function technique to analyze and measure the factors that influence the students’ achievement. This article also opened a controversial debate, not solved up to now, about the influence of the school resources in their students’ achievement. Almost all empirical analyses find that the students’ personal characteristics and their family and social environment have a decisive influence in their outcome. The controversy focuses on the variables related to the school characteristics. Some authors support the idea of a weak relationship between school resources and students’ achievement (Coleman et al. 1966; Hanushek 1986, 1997, 2003). As a reaction of this attitude, other researchers emphasize the relevance of some school resources (Summers and Wolfe 1977; Card and Krueger 1992; Hedges, Laine and Greenwald 1994; Vargelsteg...
and King 1998; Figlio 1999; Dewey 2000; Krueger 1999, 2003). This controversy shows the problems of the education production function analysis related to the available data, its relevance, and the statistical models used to measure its influence.

The literature measures the school outcome using the standardized achievement test scores. Obviously, these tests are not specifically designed to measure the future students’ performance in the labor market, and they focus only on basic achievement (usually language and mathematic skills). There is a problematic connection between the present school achievement and the future post-school outcome of these students in the labor market. That is the reason why these tests are more appropriated to some students or jobs than others, and, for example, they tend to be more suitable in the earlier grades. Another problem with these tests is that the individual test score information for each student is seldom available, but only the aggregate information for schools uses to be accessible.

Inputs have also availability problems, so to avoid the temptation of choosing only the available inputs, it is important to classify them and take into account their most important characteristics. The main inputs can be classified in two groups, 1) the socio-demographic characteristics of students’ families and peers (income, parental education), and 2) the school organization (class size, administrative expenditure) and teachers characteristics (culture, wages, tenure, experience). There is a third input related to the personal characteristics of the students that define their learning potential. Obviously, a smarter or hard-working student will have a better achievement, thought it is very difficult to measure those capacities. Moreover, all the inputs interact between them and with these innate personal abilities to produce a specific outcome. Besides, the education process is cumulative, so past inputs have some influence in present output. Though it is very difficult to measure all these variables, there is an indirect way to take them into account, by using past outputs (that is, test scores in previous years) as an input of potential output level (Bacdayan 1997; Goldhaber and Dominic 1997).

The input analysis has been concentrated in the school characteristics, trying to measure which of these features are relevant for the output, taking always into account the personal, familiar and peer characteristics. There are a lot of studies about this issues applied to different educational systems, with the problem that their empirical results are very different and problematic to compare because each study uses different variables. Some articles have made compilations of articles that compare the empirical results of different studies, again with varying results. Others have made international comparisons, either using world-wide scholastic performance tests (PISA, TIMSS), or making compilations of different articles from different countries. Once more, the results of these comparisons are varied: while some authors support the fact that influence of school characteristics in students achievement is higher in developing countries than in the developed ones (Fuller and Clark 1994; Glewwe 2002), other researchers maintain the relative irrelevance of school resources in their performance (Hanushek and Luque 2003).

The inconclusive results of the input-output analysis in the educational system show the ignorance of the shape of the education production function. Besides the problems with data availability, the schools are complex organizations that deal with complex agents (students, teachers, parents), so relationships between inputs and outputs are not easy to analyze. Just having more computers or more teachers per student does not assure that the achievement of the students will increase.

In the economic educational analysis, the main efforts to identify the contexts, processes and methodologies that transform educational inputs in students’ outcomes, have been made through an increasing sophisticated statistical analysis. If individual data of each student are available, then the researcher can use the Data Envelopment Analysis (DEA), a non-parametric analysis based on models of mathematical optimization (Worthington 2001). This technique overcome the problem of the ignorance of the production function shape (it is not necessary to specify it in the model), and is very appropriate to measure the efficiency of schools resources. But when data are aggregated and it is important to measure the influence of inputs in the school achievement, the econometric or parametric analysis is the best option. However, this second option has also limitations. One is the diminishing returns of inputs when their use is very intensive (non-significant coefficients would reflect then this congestion point but not the irrelevance of the variable). Other problem is related to data aggregation, because aggregate-level correlations have not to be the same as individual-level correlations. For example, some inputs could be more positive for advanced students and others more appropriate for backward ones (Eide and Showalter 1998). One way to overcome this problem is the use of multilevel models, because students are not simple random samples from the population of students, their performance within the same class will be correlated, as will be the performance within the same school or the same district. It must be established the level at which it is measured each of the variables.

In conclusion, the influence of specific inputs on students depends mainly on the characteristics of each educational system. Therefore, next section will analyze the Spanish educational system.
THE SPANISH EDUCATIONAL SYSTEM

The arrival of Democracy in Spain increased the public social expenditures in health, pensions, and education. There is a great interest of governments and society in evaluating the efficiency of these expenditures. Almost all studies find a negative outcome of the Spanish educational system, because the international comparisons studies (the most relevant for the Spanish case is PISA) show bad outcomes of our students in spite of the increasing resources expended in the schools (Fuentes 2009). The main problems are the comparative lower level of the Spanish students, and the higher rates of academic failure and abandon before finishing the secondary school.

The Spanish educational system has some relevant characteristics. One of them is the existence of three types of schools: public schools, controlled by the government and where students do not have to pay tuition, private schools, which have to follow only some simple and general education governmental rules, and a third group of schools, named “concertados”, which are private schools that receive government vouchers for most of the student’s tuition and imposes a commitment on the part of the school to adhere to certain rules and standards outlined by the government (students-class ratio, students-teachers ratio, academic background of teachers, syllabus, and students draft system). Another outstanding characteristic of the Spanish educational system is that, in the last years, the responsibilities in education have been transferred from the central state to the regional governments.

The decentralization of the Spanish educational system and the previous elimination of the homogeneous academic achievement tests for scholar students, explain the scarcity of individual students’ achievement information in Spain. Researchers have used mainly the University Entrance Examination, the only standardized test for all the Spanish students made at the end of the secondary school (San Segundo 1991; Pedraja and Salinas 1996; Mancebón 1998; Mancebón and Bandrés 1999; Muñiz 2000; Seijas 2005; Dios-Palomares et al. 2006), the PISA data on Spanish students (Santin 2006; Calero and Escardíbulo 2007), and other indicators when available, like specific tests made in schools by the regional governments (Cordero et al. 2005). All these studies use the DEA analysis to measure the efficiency of the schools. In most of the cases, the sample is a limited number of schools in one specific city or region. All these studies find a close relationship between the student’s socioeconomic level and their performance.

Other studies use grouped data and parametric models to investigate some specific characteristics of the Spanish educational system, like the comparison between public and private schools (Calero and Escardíbulo 2003; Sancho 2008) and the role of immigrants in the educational system (Salinas and Santin 2007). Both factors are related, given that the concentration of immigrant students in public schools can reduce their achievement.

The Regional Government of Madrid has developed a special test to evaluate the knowledge and capabilities in Mathematics and Language of students in their last year of primary school (12 years old). This test started in 2005 and is made in all the schools of the region. The publication of the results of these tests, grouped by schools, started a controversy because some political parties and labor unions did not like the comparisons between public and private schools that could be derived from the published results. In the next two years the results of the test were not published, and in 2008 were published again. The only study about this test, the one published in 2005, found an important influence of the economic characteristics of the student’s district or neighborhood, and the immigration through their massive presence in the public schools in some districts, although it only studied the results of the schools in Madrid city, but not in the rest of the region (Crespo et al. 2006).

Our study uses the academic results of this test in 2008, with 729 schools of the Madrid municipalities, without including schools from Madrid City. The objective is to analyze the main factors that explain the school performance at the municipality level, through not only traditional variables as the socioeconomic context of the student and the school resources, but other variables related to the importance of the municipality where the student live, and some other factors that are in the special context of the regional educational system such as the ownership of the school and the number of immigrants at the school.

METHODOLOGY

By developing a cross-section data model, this paper tries to explain –on a school-by-school basis– the potential variables which can explain the variability of the exam results of the sixth grade of primary education. Cross-section models can be used to analyze the data of a broad sample of individuals, families, enterprises, countries –or, in this specific case, the primary schools in Madrid– in a given year. Therefore, we do not have a temporal dimension.

While the time series analysis must adopt a broad set of econometric tools which are also used in the cross-section analysis, the first one is more complicated because of the usual existence of trends, correlations (and autocorrelations), dynamics, or stationarity, in economic time series.

On the other hand, although some econometric models derive from formal economic models and theories, other
studies are based on non-formal economic reasoning and the intuition. It is the case of the present paper.

Some of the most frequent problems that are observed in the cross-section data analysis are heteroskedasticity, multicollinearity, endogeneity or stochastic regressors. As can be proved, the model developed in this paper, and the results derived from it, are robust enough to the previous problems.

THE MODEL

The present study is based on the 2008 test of the sixth grade of primary education in the schools of the Madrid region. The developed model tries to explain the variability of the results of the examination of the sixth grade of primary education as a result of a set of variables. These variables can be classified in two broad groups:

1) Particular characteristics of each of the schools: available educational resources, students composition (immigrants versus domestic students), and ownership of the educational centers. Under this group we include the following variables per school:
- The student-teacher ratio (stud/teacher)
- The student-computer ratio (stud/computer)
- The foreign students (immigrants)/total students ratio (foreign/total)
- A dummy-variable (0,1,2) which reflects the titularity of the educational centers (titul): public schools, private schools that receive public funds (“concertados” schools), and private schools, respectively.

2) Economic, geographical and political characteristics of the municipality where the schools are placed: per capita income (Yp), public spending on education over the total public spending of the municipality (edspending), municipality’s distance to Madrid City (dist), and two dummy variables which represent the political party in power: PP (that adopts the value of 1 under a Popular Party government and 0 otherwise) and PSOE (which adopts the value of 1 under a Socialist Party government and 0 otherwise).

The econometric specification of the model is the following:

\[
\log(\text{grade}) = c + \beta_1(\text{stud} / \text{teacher}) + \\
\beta_2(\text{foreign} / \text{total}) + \beta_3(\text{stud} / \text{computer}) + \\
\beta_4(\text{titul}) + \beta_5(\text{Yp}) + \beta_6(\text{edspending}) + \\
\beta_7(\text{dist}) + \beta_8(\text{PP}) + \beta_9(\text{PSOE}) + \epsilon
\]  

(1)

The empirical analysis is based on the "Knowledge and Capabilities of Mathematics and Language" test developed by the Madrid Autonomous Community in the last four years.

We have used the last test made in the course 2007/2008 to 1,270 schools, 541 within the City of Madrid, and 729 in the rest of the municipalities of Madrid Region. The first objective is to compare the municipalities’ differences, so we have to erase the schools of Madrid City because it introduces distortions in the analysis. The option of dividing the Madrid’s schools into districts is not possible because there are no so many data at this level as at the municipality level. Given the lack of some relevant data at municipal level, the final analyzed sample includes 617 schools.

The data set at the municipality level comes from the National Statistics Institute (INE) and the Municipal Database of the Statistics Institute of the Community of Madrid (ALMUDENA): population age, foreign population, uneducated population, distance to Madrid City, political party governing the municipality, and per capita income level. Education expenditure of each municipality (data from the Minister of Finance) is also included.

The competence in education expenditure is at the regional level (the government of the Community of Madrid), even though each municipality can add expenditures to expand the education effort in its municipality. We have used education expenditure as a percentage of the total municipal budget.

Besides the municipality level data, we have introduced other data at a school level: number of students and foreign students, number of teachers, classes, and computers. All these data are provided by the Education Services of the Autonomous Community of Madrid.

Table 1 shows the a priori expected signs for each of the previous variables. On the other hand, Table 2 reports the results of the estimation of equation (1) using WLS.

| Students/teacher ratio (stud/teacher) | (+) |
| Foreign students/total students (foreign/total) | (+) or (-) |
| Students/computer ratio (stud/computer) | (+) |
| Titularity (titul) | (+) or (-) |
| Per capita income (Yp) | (+) |
| Education spending/total spending (edspending) | (+) |
| Distance from the capital (dist) | (+) |
| Popular Party government (PP) | (+) or (-) |
| Socialist Party government (PSOE) | (+) or (-) |

Table 1. Expected signs for the model regressors.
Dependent Variable: log(grade)
Method: Weighted Least Squares
Sample: 1 617
Included observations: 617
Weighting series: 1/\(\log(Y)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2.88500</td>
<td>28.07290</td>
<td>0.0000</td>
</tr>
<tr>
<td>student/teacher</td>
<td>-0.001937</td>
<td>-2.296731</td>
<td>0.0238</td>
</tr>
<tr>
<td>foreign/total</td>
<td>-0.264929</td>
<td>-6.10887</td>
<td>0.0000</td>
</tr>
<tr>
<td>student/computer</td>
<td>-0.022902</td>
<td>-2.049431</td>
<td>0.0406</td>
</tr>
<tr>
<td>titul</td>
<td>0.107168</td>
<td>3.749950</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
| \(\log(Y)
| 1.216000     | 3.587913    | 0.0004|
| edspending       | 0.424601     | 1.924220    | 0.0548|
| dist             | -0.002171    | -3.795475   | 0.0002|
| PP               | 0.017862     | 0.787789    | 0.4311|
| PSOE             | -0.008827    | -0.377415   | 0.7080|

Table 2. Regression results.

All variables, except the dummy variables for the political party governing the municipality (PP and PSOE), show the expected sign and are significant at the 5% level. It is very relevant to underline the very high significance of variables like the foreign students/total students ratio, the ownership of the school (public-private), the per capita income level of the municipality, and the municipality’s distance to Madrid City. Finally, the weighted fit of the model is relatively good (\(R^2=0.64\))

A simple reading of Table 2 shows that:

- The higher the student-teacher ratio,
- The higher the student-computer ratio,
- The higher the foreign students/total students ratio,
- The lower the per capita income level of the municipality,
- The lower the education public spending/total municipal budget ratio,
- The greater the municipality’s distance to the capital...

...the lower is the average mark for the exam.

On the other hand, it is relevant to emphasize on the existence of a positive and significant relationship between private “concertados” schools and the average mark of the examination.

Therefore the estimated model is as follows:

\[
\text{log(grade)} = 2.8885 - 0.0019(\text{student/teacher}) + \\
-0.2648(\text{foreign/total}) - 0.0229(\text{student/computer}) + \\
+0.107(titul) + 0.4246(\text{edspending}) + \\
-0.0021(\text{dist}) + \varepsilon,
\]

Figures 1 and 2 demonstrate that model residuals do not show any obvious patterns that seem inconsistent with the assumptions of the model.

CONCLUSIONS

This article develops a model that explains the variability of the examination marks of the sixth grade of primary education as a result of two broad set of variables: 1) particular characteristics of each school (available educational resources, students composition, and ownership of the educational centers), and 2) economic, geographical and political characteristics of the municipality where the schools are placed (per capita income level, public spending on education over the total public spending of the municipality, distance to
Madrid City, and the influence of the political party in power).

The cross-section data model is estimated under a Weighted Least Squares (WLS) method for correcting heteroskedasticity, and the model demonstrate that the examination results of the sixth grade of primary education in the schools of Madrid region are negatively related to the available educational resources of each of the primary schools individually considered (student-teacher ratio and student-computer ratio). On the other hand, there is a negative and significant relationship between the proportion of foreign students in the total number of students and the examination grades. The third point to be considered is the existence of a positive and significant relation between private/“concertados” schools and the results of the examination.

Additionally, it must be emphasized the existence of a clear and significant positive relationship between municipalities’ economic factors (per capita income level and public spending on education over total public spending) and the results of the exam.

Finally, we can define as a non-expert result the existence of statistically significant relationship between the municipalities’ distance to Madrid City and the results reached by the primary school students. Is it maybe the proximity to the capital a factor of major and better access to education and culture for the students of primary schools in the Community of Madrid?

**APPENDIX**

<table>
<thead>
<tr>
<th>ALUMEXT</th>
<th>ALUMORD</th>
<th>ALUMPROF</th>
<th>TITUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>-0.032</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>-0.315</td>
<td>-0.029</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>-0.458</td>
<td>-0.132</td>
<td>0.443</td>
<td>1.000</td>
</tr>
<tr>
<td>-0.152</td>
<td>-0.030</td>
<td>0.180</td>
<td>0.320</td>
</tr>
<tr>
<td>-0.037</td>
<td>0.031</td>
<td>-0.049</td>
<td>-0.191</td>
</tr>
<tr>
<td>0.072</td>
<td>-0.027</td>
<td>-0.019</td>
<td>-0.113</td>
</tr>
<tr>
<td>-0.031</td>
<td>0.007</td>
<td>0.042</td>
<td>0.066</td>
</tr>
<tr>
<td>0.072</td>
<td>-0.015</td>
<td>-0.080</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOGYP1</th>
<th>GASTOED</th>
<th>DIST</th>
<th>PP</th>
<th>PSOE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.070</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Correlation matrix of model regressors

**REFERENCES**


MIGUEL ÁNGEL ALONSO NEIRA is full-time professor in Applied Economics at the Rey Juan Carlos University of Madrid. Graduate in Economics and Business at the Complutense University of Madrid (1993), he studies a MSc in Monetary Economics and International Finance at the University of Essex (September 1999) with a scholarship of “La Caixa” and the British Council. After finishing the Master’s Degree in September 2000, he is awarded the Leatherland Boook Prize to the best postgraduate student of this University. PhD in Economics by the Rey Juan Carlos University of Madrid (December 2001), in December 2003 he is awarded the “Victor Mendoza” National Prize for his research “Capital Controls, Financial Variables and Economic Policy”. His e-mail address is: miguelangel.alonso@urjc.es.

LUIZ PIRES JIMENEZ is Professor and Assistant Dean at the University Rey Juan Carlos in Madrid. Graduate in Economics and Business at the Complutense University of Madrid (1992) and PhD in the same University (1999). His e-mail address is: luis.pires@urjc.es, and his Web-page can be found at http://www.luispires.weebly.com/.

ALBERTO ROMERO is Master in Business Administration, Master in International Economic Analysis and Graduated with an award for Excellence in the five years Degree Business Administration and Management at Rey Juan Carlos University. He worked at the International Telecommunication Union (United Nations) in Geneva and nowadays he is Lecturer at the Applied Economy Department at Rey Juan Carlos University. In Austria he specialized studying an English program in Export-Oriented Management at the University of Applied Sciences of Krems. He got two grants to participate in two programs about ethnology, declared of universal interest by UNESCO and coordinated by the Complutense University of Madrid. His e-mail address is: alberto.romero@urjc.es, and his Web-page can be found at http://spanisheconomy.weebly.com/.

AUTHOR BIOGRAPHIES

ROCÍO GALLEGO LOSADA is professor in Financial Economy and Accounting at the Rey Juan Carlos University of Madrid. Graduate in Economics and Business Administration at the Autonoma University of Madrid (1992), she has been a visiting scholar in European Universities in Germany (2004), France (2006), Portugal (2008) and in the University of Nevada, USA (2008). After finishing her PHD (2000), she received a grant of “Mapfre” to research about Pension Plans. Her e-mail address is: rocio.gallego@urjc.es


