Universidad
Carlos III de Madrid

PISA 2009: AN ANALYSIS OF THE DIFFERENCES IN TEST SCORES ACROSS REGIONS IN SPAIN

JUAN GUERRA GARCIA (100083756)

TUTOR: PROF. DR. D. CARLOS SAN JUAN MESONADA

JUNE 2012

ABSTRACT:

Since the implementation of the Programme for International Student Assessment (PISA) in the year 2000, Spain has had results below the Organisation for Economic Cooperation and Development (OECD) average. Data shows there are big differences in PISA results across some of the regions in Spain. The analysis we have performed proves family background and individual characteristics have an important effect on the student's achievements and are part of the explanation for differences across regions. However, after controlling for these two factors, there is still statistically significant evidence for the difference in test scores between Madrid and some of the regions in the South. An innovative historical approach has been followed that incorporates literacy rates in 1900 to the analysis across regions. Results show a strong correlation between literacy rates at the beginning of the twentieth century and PISA test scores 110 years later. The correlation could be due to an intrinsic cultural component in individuals, affecting their incentives and preferences; interpreting literacy rates as a proxy for social institutions in 1900 could lead to another explanation, or it could be explained through the land inequality hypothesis and its impact on the accumulation of human capital.

TABLE OF CONTENTS:

INTRODUCTION:

In the year 2000, the OECD implemented the Programme for International Student Assessment in order to evaluate the skills and knowledge of 15-year-old students and test the education levels across the different countries. These tests focus on three main areas: mathematics, reading and science, and students take them close to the completion of compulsory education in order to have an indicator of their level of knowledge prior to their complete immersion in society.

Since then, every three years (2000, 2003, 2006 and 2009), more than 70 countries have analyzed through PISA their students' results and used them as a tool for developing policies with the aim to improve education outputs. Since the year 2000, Spain has had bad results, below the OECD average. These poor results are not homogeneous and general across the whole country; Spain's low performance is even more accentuated depending on the region.

PISA's study measures education outputs; it is an evaluation of student's cognitive skills and, bearing in mind Hanushek and Woessmann's (2008)[i] statement "*cognitive skills independently affect economic outcomes even after allowing for economic institutions, property rights…*", an insurance for the long-term development of a region. There are statistically significant differences across regions, and a clear polarization North-South can be clearly differentiated.

Many variables have an impact on student's achievements, but what appear to be two determinant factors are family background and individual characteristics. The first objective of this project is to quantify the impact of these variables, control for these two factors, and evaluate again the differences across regions in Spain. Previous literature finds evidence on the correlation between the parent's education and their descendants' performance. Davis-Kean (2005)[ii] concludes that parental education achievements are tightly related to their children's school scores, and Haveman and Wolfe (1995)[iii] also confirm that children of more educated parents have a greater success in school. The first part of this project estimates a model on the influence of individual and family background characteristics on the education achievements of children. The model is used to ask the following question: How would the differences across regions be if these two factors were controlled? The question could be interpreted as… if all of the regions had the same family background, similar individual characteristics and similar household environment, would these differences change?

The second part of the project consists of an analysis of these differences. The objective is to analyze the original micro data including as well two macro perspective variables at a regional level: Gross Domestic Product (GDP) per capita and literacy rates in 1900. The latter will offer a historical innovative approach to possible explanations for the aggregate and long term determinants of these differences across regions in Spain. Motivation came from the papers by Acemoglu, Johnson and Robinson (2002)[iv], which relate historical institutions to long run prosperity and the hypothesis by Engerman and Sokoloff (1997)[v] on South America's poor growth due to high land inequality in the past. The idea is, if once we have controlled for parents' education level[vi] and part of the individuals' characteristics, can literacy rates in 1900 provide an explanation to PISA 2009 differences across regions? If there is significant evidence of this effect, how can we interpret it?

Up to my knowledge, there is no evidence in the literature of the existence of a correlation between literacy rates in the beginning of the twentieth century and PISA results 110 years later. This academic project provides first evidence and tries to generate further discussion on the subject.

The paper has a very accessible structure: first, an overview of the descriptive statistics with brief explanations of the variables used in the analysis; then, a technical approach of the analysis, regression equations and statistical and econometric software used. Following, a deep study of the results obtained, with graphs and tables to support the arguments. Finally, a conclusion that recapitulates the most important aspects of the project and shows the completion of the initial objectives.

VARIABLES & DESCRIPTIVE STATISTICS:

The analysis is done using PISA 2009 database. It is a very big and complex database composed of hundreds of questions and information on thousands of students across 65 countries worldwide. For the study, object of our discussion, the data was filtered and only the Spanish data was selected. The final data set contained information on 23708 Spanish students across different regions (although not every Autonomous Community in Spain decided to take part in PISA 2009). Table 1 shows the average test score in math and reading in each of the regions for which data is available. They are ranked from lowest to highest average grade in math. Differences across some of the

regions are very big, even more bearing in mind that PISA considers a gap of 39 points like the difference of a formal school year[vii].

Table 1: Average PISA scores in math and reading

REGION	MATHEMATICS	READING
ISLAS CANARIAS	434.94	448.02
ANDALUCIA	461.71	460.5
ISLAS BALEARES	464.47	457.26
MURCIA	477.91	480.1
GALICIA	489.18	485.53
ASTURIAS	493.58	490.22
CANTABRIA	494.65	487.85
CATALUÑA	495.64	498.02
MADRID	496.48	503.47
LA RIOJA	503.54	497.91
ARAGON	505.6	495.33
PAIS VASCO	509.65	494.46
NAVARRA	511.24	497
CASTILLA Y LEON	514.33	503.04

In order to control the different individual and family background characteristics of each student, several variables have been selected. On the one hand, individual characteristics, like *SEX* (dummy = 1 if student is a male), *ST_INMIGRANT* (dummy = 1 if student is an immigrant), *PRE_PRIMARY* (dummy =1 if student attended pre primary education), SING_PARENT (dummy = 1 if student lives in a household with just one of his/her parents). On the other hand, variables to control family background and the studying environment at home. Parents' education was measured according to the highest level of schooling achieved by the parent with the highest degree. According to the International Standard Classification of Education (ISCED), six different levels of education can be distinguished; in this study they will be grouped in four: ISCED *1 or LESS* (Parents who have at most primary education), *ISCED 2* (parents who have completed lower secondary education), *ISCED 3 or 4* (parents who have finished upper secondary education or post secondary non tertiary education) and *ISCED 5 or 6* (parents who have completed tertiary education). It is clear from Table 2 the differences in the average grade in math for students depending on their parent's education. Parents' education is on average a big constraint; there is a lack of intergenerational mobility and big inequalities in educational opportunities or at least inequality in educational outcomes, but this could be the subject of another discussion.

Table 2: Average grade in math for the whole sample of Spanish population depending on parents ISCED level

REGION	ISCED 1 OR LESS	ISCED 2	ISCED 3 OR 4	ISCED 5 OR 6
AVERAGE GRADE IN MATH	434	462	481	509

Part of the differences across regions will eventually be explained by the differences in parents' education. Table 3 presents the proportion of parents who have achieved at most each of the ISCED levels in the different regions. The distribution of levels across regions is very heterogeneous and a key factor for the variance in results across the autonomous communities.

Table 3: Proportion of parents who have achieved at most each of the following ISCED levels

REGION	ISCED 1 OR LESS	ISCED 2	ISCED 3 OR 4	ISCED 5 OR 6
ANDALUCIA	0.15	0.24	0.25	0.36
ISLAS CANARIAS	0.14	0.20	0.30	0.36
MURCIA	0.12	0.22	0.28	0.38
ISLAS BALEARES	0.09	0.20	0.26	0.45
GALICIA	0.09	0.19	0.28	0.44
CATALUÑA	0.07	0.17	0.26	0.50
ARAGON	0.06	0.16	0.28	0.50
LA RIOJA	0.06	0.17	0.29	0.48
MADRID	0.05	0.17	0.28	0.50
NAVARRA	0.05	0.13	0.28	0.54
CASTILLA Y LEON	0.05	0.17	0.30	0.48
PAIS VASCO	0.04	0.09	0.25	0.62
ASTURIAS	0.04	0.16	0.30	0.50
CANTABRIA	0.04	0.16	0.30	0.50

Number of books at home is a proxy for the social and educational environment in the student's household, the importance of this variable is highlighted by Ludger Woessman[viii]. In this study, two dummies were created in order to control for the two extremes of the population *L*ESS_25BOOKS (dummy = 1 if there are less than 25 books in the student's home) and MORE_200BOOKS (dummy = 1 if there are more than 200 books in the student's home). Two more variables complete the micro variables used in the analysis, *DESK* (dummy = 1 if the student has a desk to study at home) and *COMPUTER* (dummy = 1 if the student has a computer in his house). The selection of these variables has been made after a literature review; selecting those variables that will provide a better control for individual characteristics and family

background. In the last part of the project, two more variables are included, GDP_PER_CAPITA (in thousands of Euros, data obtained for the year 2009 from the Spanish National Institute of Statistics (INE)) and LIT_RATE_1900, the literacy rate in each region in 1900. Literacy rates 100 years ago were obtained from Carlos Barciela López and Albert Carreras' book[ix] and contrasted in other sources[x]. LIT_RATE_1900 is a proxy for the development of social institutions at the beginning of the 20th century, it is a measure for initial differences 100 years ago and might prove that even after controlling for individual characteristics, family background and GDP per capita, there is something else, either cultural or institutional, which produced the inception of those initial differences which even in the year 2009 still persist.

METHODOLOGY:

Handling PISA 2009 is not an easy task. In order to measure school outputs, PISA uses plausible values (PV), instead of using just a single measure of the students' score. These plausible values are "random numbers drawn from the distribution of scores that could be reasonably assigned to each individual – that is the marginal posterior distribution"[xi]. For each student and subject, five PV's are reported; it is necessary to highlight the importance of handling the data with PISA[xii] recommendations; in order to calculate unbiased estimates it was necessary to run each regression model five times. The approach suggested by PISA also takes into account 80 replicate sampling weights for each student, in order to obtain unbiased estimates of the standard errors (SE). These weights are taken into consideration because not every student is equally representative of the population in each region, even though they were randomly selected. To do so it was necessary to use the statistical software SPSS, and a macro provided by PISA (MCR_SE_REG), which after executing the command is applied sequentially computing more than 400 coefficients per independent variable, which after being treated as follows in the PISA documentation produces unbiased estimates and unbiased standard errors. Not including the 80 student replicate weights could lead to underestimated SE's, which might result in committing an Error Type I (reporting the estimate as statistically significant when in fact this is not true), the estimates would give information of the students in the sample but not of the whole population of the region.

We will divide our study in four regression models, to which we will refer as equations I, II, III and IV. For each equation we will analyze the results in math and/or reading.

With the first equation it will be proved that there exist statistically significant differences across regions within Spain. In order to do the analysis a comparison of the results across regions will be done taking as a benchmark the results in Madrid (β_0).

$$MATHGRADE = \beta_0 + \beta_1 ANDALUCIA + \beta_2 ARAGON + \beta_3 ASTURIAS + \beta_4 CANTABRIA + \beta_5 CASTILLA\_Y\_LEON + \beta_6 CATALUNA + \beta_7 GALICIA + \beta_8 ISLAS\_BALEARES + \beta_9 ISLAS\_CANARIAS + \beta_{10} LA\_RIOJA + \beta_{11} MURCIA + \beta_{12} NAVARRA + \beta_{11} PAIS\_VASCO + \varepsilon \quad \text{(I)}$$

In the second equation a small Human Capital model will be build, trying to control for two main factors: individual characteristics and household environment, as we have seen in descriptive statistics. The first two equations could be taken as granted, but in order to do our analysis more precise we will include part of the results in the project.

$$MATHGRADE = \beta_0 + \beta_1 SEX + \beta_2 INMIGRANT + \beta_3 PREPRIMARY + \beta_4 SINGPARENT + \beta_5 LESS25BOOKS + \beta_6 MORE200BOOKS + \beta_7 DESK + \beta_8 COMPUTER + \beta_9 ISCEDLESS1 + \beta_{10} ISCED2 + \beta_{11} ISCEDMORE5 + \varepsilon \quad \text{(II)}$$

The third regression is a combination of equations (I) and (II), the objective is to see the differences across regions once we have controlled for individual and household characteristics. It could be interpreted as, what would happen if all the families in the different regions where the same, would we still have those differences?

$$MATHGRADE = \beta_0 + \beta_1 SEX + \beta_2 INMIGRANT + \beta_3 PREPRIMARY + \beta_4 SINGPARENT + \beta_5 LESS25BOOKS + \beta_6 MORE200BOOKS + \beta_7 DESK + \beta_8 COMPUTER + \beta_9 ISCEDLESS1 + \beta_{10} ISCED2 + \beta_{11} ISCEDMORE5 + \beta_{12} ANDALUCIA + \beta_{13} ARAGON + \beta_{14} ASTURIAS + \beta_{15} CANTABRIA + \beta_{16} CASTILLA_{Y_{LEON}} + \beta_{17} CATALUNA + \beta_{18} GALICIA + \beta_{19} ISLAS_{BALEARES} + \beta_{20} ISLAS_{CANARIAS} + \beta_{21} LA_{RIOJA} + \beta_{22} MURCIA + \beta_{23} NAVARRA + \beta_{24} PAIS_{VASCO} + \varepsilon \quad \text{(III)}$$

The last regression will lead us to the final discussion of the paper: do literacy rates in 1900 have an effect in PISA results 2009?

$$MATHGRADE = \beta_0 + \beta_1 SEX + \beta_2 INMIGRANT + \beta_3 PREPRIMARY + \beta_4 SINGPARENT + \beta_5 LESS25BOOKS + \beta_6 MORE200BOOKS + \beta_7 DESK + \beta_8 COMPUTER + \beta_9 ISCEDLESS1 + \beta_{10} ISCED2 + \beta_{11} ISCEDMORE5 + \beta_{12} GDP\_PER\_CAPITA + \beta_{13} LIT\_RATE\_1900 + \varepsilon \quad \text{(IV)}$$

ANALYSIS & RESULTS:

In Table 1 we saw the ranking of the average test score in the different regions in mathematics and reading. It could be clearly stated the big variance of results. Which differences are statistically significant?

In order to make it easier to interpret the results, the rest of the analysis will be done in comparison to Madrid. This region has been chosen as a benchmark for many reasons; amongst them, it is the capital of the country object of our study and for further analysis north vs. south it is located in the centre of the country. The analysis will be divided in three blocks: block 1, overview of the initial differences across regions and the effect and importance of the variables selected; block 2, control for individual characteristics and family background, persistence of these differences?; block 3 focuses on possible explanations for the differences across regions and tests the hypothesis of a correlation between literacy rates in 1900 and PISA results 2009.

BLOCK 1: Table 4 compares the differences in mathematics in each of the regions, which decided to take part in PISA 2009, with Madrid. The regions with an asterisk (*) are significant at a 5% level. Pais Vasco, Navarra and Castilla y Leon's students obtain, on average, better grades than Madrid, on the contrary, Andalucia, Islas Baleares, Islas Canarias and Murcia obtain significantly worse results. There is no significant evidence for the differences with Aragon, Asturias, Cantabria, Cataluña, Galicia and La Rioja.

Table 4: Differences across regions in mathematics (in comparison to Madrid)

REGION	COEFF.	SE	T-STATISTIC
CONST	496.48	4.38	113.35
ANDALUCIA *	-34.77	6.65	-5.23
ARAGON	9.12	6.87	1.33
ASTURIAS	-2.9	6.31	-0.46
CANTABRIA	-1.83	6.64	-0.28
CASTILLA Y LEON *	17.85	5.89	3.03
CATALUÑA	-0.84	7.35	-0.11
GALICIA	-7.3	6.13	-1.19
ISLAS BALEARES *	-32.01	4.84	-6.61
ISLAS CANARIAS *	-61.54	5.94	-10.36
LA RIOJA	7.06	4.94	1.43
MURCIA *	-18.57	7.16	-2.59
NAVARRA *	14.76	5.77	2.56
PAIS VASCO *	13.17	5.15	2.56

Can equation II provide part of the explanation for these differences? Table 5 shows that most of the variables object of our study have a significant effect. If we take a look at the individual characteristics, male students will obtain on average nearly 21 more points in math and, on average, immigrants and children from single parenting families obtain worse results, -28 and -11 respectively. For the sake of briefness, data for only one of the subjects will be presented; coefficients are very similar independently of the subject. As a quick remark, I would like to highlight that while boys obtain on average 20.85 points more than girls in math, on reading tests the statistics are the exact opposite, girls obtain on average 25.13 more. The attendance of the students to pre-primary education has a very strong and positive effect; numerous studies highlight the importance of pre-school education even in the long run, having an impact on graduation from high school rates and income.

It can be confirmed that parent's education has a very important effect on their children's scores. Davis-Kean[ii] and Haveman and Wolfe's[iii] results have been validated and show similar results to regression II: the higher the level of education of the parents the greater the success of their children in school. There is a difference of 40 points between students with parents who have completed tertiary education and parents with no studies. The variables used to analyze the effect of the environment of study in each household have the expected significant effect. Those students who have access to their own desk to study, and their families have an interest for culture and literature, obtain higher grades. Having or not a computer doesn't seem to have a significant effect; this might be due to the double impact the possession of a PC may have on a student: on the one hand, it can be a very good source of knowledge and information; on the other hand, it is a distraction if used for games or social networks. The strong and negative effect the variable LESS_25BOOKS has on the students test score, -82.73 on average points less than those students who stated on the PISA questionnaire had at home more than 200 books, needs to be remarked. This result is even more impressive considering that PISA considers a gap of 39 points like the difference of a formal school year, as stated before.

Overall, Table 5 shows the importance of the variables chosen as controls of individual and family background characteristics. The regression shows these variables are an important explanation for the variance of results across individuals.

Table 5: Marginal effect of the variables chosen on mathematics test scores

	COEFF.	SE	T-STATISTIC
CONST	442.43	9.02	49.05
COMPUTER	4.16	8.37	0.50
DESK *	22.98	4.29	5.36
LESS_25BOOKS *	-52.91	2.61	-20.27
MORE_200BOOKS *	29.82	2.18	13.68
PARISCED_2 *	-11.98	3.01	-3.98
PARISCED_LESS1 *	-27.21	4.74	-5.74
PARISCED_MORE5 *	12.91	2.55	5.06
PRE_PRIMARY *	17.46	5.32	3.28
SEX *	20.85	1.82	11.46
SING_PARENT *	-11.3	3.04	-3.72
ST_INMIGRANT *	-28.44	3.06	-9.29

BLOCK 2: In block 1, we confirmed the existence of differences across regions, and that the variables selected for the human capital model were statistically significant. The objective of this second block is to merge equations I and II; in other words, analyse the differences across regions once we have controlled for family background and individual characteristics.

Table 6 shows the result of running the regression in equation III. First of all, the variables controlling for family background and individual characteristics have similar coefficients and are all (except having a PC at home) statistically significant at a 5% level. What has happened to the differences across regions?

Results obtained are very interesting. If we take a look at Figure 1, we can see in a very visual way how, after controlling for individual characteristics and family background, the differences in test scores in math (attached in the appendix is the graph for reading, Figure 2) are considerably reduced. On the horizontal axis are the conditional differences of the regions with respect to the average in Spain; on the vertical axis are the non-conditional differences. Two conclusions can be obtained from Figure 1; on the one hand, once we have controlled for differences in individual characteristics and family background, differences converge and are reduced; on the other hand, there are still considerable differences across some of the regions. In order to facilitate the comprehension of Figure 1, some accessory lines have been included in the figure, highlighting differences across regions.

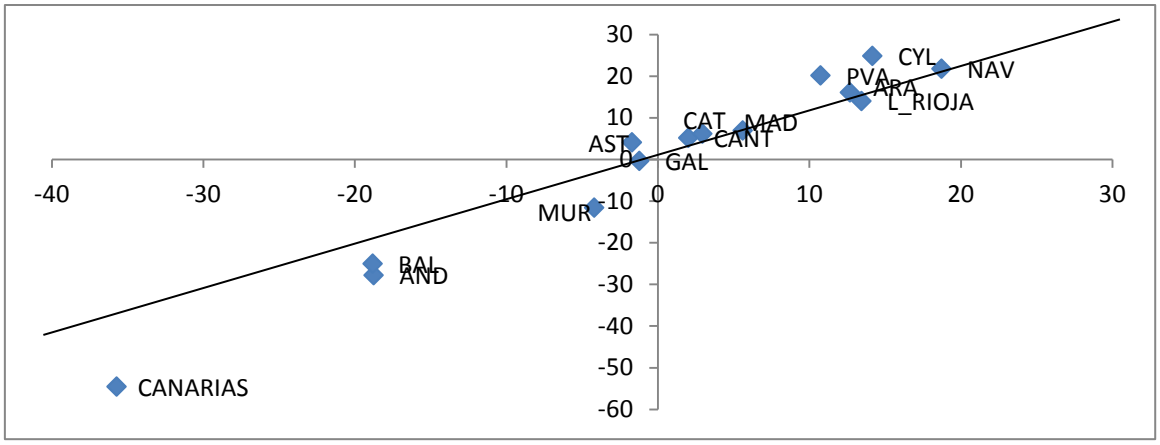

Figure 1: Conditional and unconditional differences in relation to the mean in math

T-statistics in Table 6 provide a lot of information. Again we have the results in comparison to Madrid, but now after controlling for individual characteristics and family background. What has happened to the differences across regions in comparison to Madrid? Castilla y Leon, Pais Vasco and Murcia are no longer significant at a 5% level; only Andalucia, Islas Canarias, Islas Baleares and Navarra have significant differences with Madrid. Navarra is a completely different case to the first three regions, because Navarra has a statistically significant difference in a positive way, they have the best results in the country even after controlling for numerous characteristics.

Table 6: Control for individual and family background characteristics, effect on differences across regions

	COEFF.	SE	T-STATISTIC
CONST	449.17	9.13	49.20
COMPUTER	4.61	8.46	0.54
DESK *	22.46	4.33	5.19
LESS_25BOOKS *	-48.83	2.46	-19.85
MORE_200BOOKS *	27.36	2.12	12.91
PARISCED_2 *	-9.81	3.03	-3.24
PARISCED_LESS1 *	-22.53	4.55	-4.95
PARISCED_MORE5 *	12.13	2.49	4.87
PRE_PRIMARY *	19.66	5.53	3.56
SEXO *	21.25	1.79	11.87
SING_PARENT *	-10.58	2.91	-3.64
ST_INMIGRANT *	-31.92	3.11	-10.26
ANDALUCIA *	-24.35	4.93	-4.94
ARAGON	7.07	5.86	1.21
ASTURIAS	-7.3	5.44	-1.34
CANTABRIA	-3.59	5.29	-0.68
CASTILLA Y LEON	8.56	4.69	1.83
CATALUÑA	-2.67	5.33	-0.50
GALICIA	-6.82	5.01	-1.36
ISLAS BALEARES *	-24.43	4.08	-5.99
ISLAS CANARIAS *	-41.32	4.88	-8.47

LA RIOJA	7.84	4.28	1.83
MURCIA	-9.81	5.32	-1.84
NAVARRA *	13.13	4.64	2.83
PAIS VASCO	5.14	3.84	1.34

Summing up the results in block 2, part of the differences across regions can be explained by a vector of individual characteristics (sex, pre-primary, immigrant, family structure), family background and household environment features (parents' education, environment at home, …). If all the families in Spain had similar characteristics in relation to Madrid there would be no significant differences with Castilla y Leon, Cantabria, Asturias, Aragon, Cataluña, Galicia, La Rioja or Pais Vasco; only Navarra would be doing outstandingly well. On the other hand, there are some regions that even after controlling for differences between households, still have very big differences; these are Andalucia, Islas Canarias, Islas Baleares and even Murcia at a 10% significance level. There is a clear polarization north vs. south in Spain and the objective of the next block is to introduce two more variables that might lead to an explanation for these differences.

BLOCK 3: The polarization north vs. south in Spain is evident, even after controlling for individual characteristics and family background there is still evidence of differences across regions in the north and regions in the south. The initial hypothesis that could be argued to explain part of the remaining differences in education outputs across regions is the differences in GDP per capita. The output in regression IV shows that the initial hypothesis cannot be rejected at a 10% significance level, although it can at a 5%. It can be interpreted that GDP_PER_CAPITA does have a significant positive impact but it is not as strong as it could be initially thought.

Table 7: Literacy rates in 1900 and GDP per capita effect on PISA 2009

	COEFF.	SE	T-STATISTIC
CONST	400.54	13.49	29.69
COMPUTER	4.16	8.31	0.50
DESK *	22.55	4.34	5.20
LESS_25BOOKS *	-50.5	2.52	-20.04
MORE_200BOOKS *	28.07	2.09	13.43
PARISCED_2 *	-10.23	3.06	-3.34
PARISCED_LESS1 *	-23.31	4.65	-5.01
PARISCED_MORE5 *	12.12	2.47	4.91
PRE_PRIMARY *	19.75	5.45	3.62
SEXO *	21.12	1.79	11.80
SING_PARENT *	-11.54	2.92	-3.95
ST_INMIGRANT *	-32.25	3.19	-10.11

LIT_RATE_1900 *	40.23	13.64	2.95
GDP_PER_CAPITA *	0.90	0.55	1.66

The other variable object of our study is literacy rates in 1900. As can be seen in Table 7 it has a significant and positive impact (t-statistic > 2) at a 5% significance level. The interpretation of the coefficient for LIT_RATE_1900 (40.23) is that an increase of 0.1 (10% of the population) in the literacy rate in 1900 across the different regions translates into an increase of nearly 4 points in PISA scores. The estimate calculated by the regression should not be taken into consideration in such a precise way but in a more general way; it can therefore be concluded that the initial hypothesis is correct.

Literacy rates in 1900 provide an explanation for differences across regions in PISA results 2009, how can it be that these two variables are correlated?

Acemoglu, Johnson and Robinson (2002)[iv] found evidence on the tight relationship between historical institutions and long run prosperity, and Engerman and Sokoloff (1997)[v] on the effect of land inequality on South America's poor growth. Can we extrapolate these explanations to explain the initial differences across regions in Spain?

Initial differences across regions in Spain might be one of the possible explanations for the diverse levels of literacy rates in 1900 in the regions in Spain that, at least through education outputs, persist one hundred years later. Rafael Dominguez Martín has made some research on similar subjects and states[xiii] "inequalities within land distribution, and its impact in the long run on the accumulation of human capital, would be the key determinants of underdevelopment in some regions". But how can these differences in land distribution have an impact on the accumulation of human capital? A possible hypothesis is described by Anandi Mani's model[xiv], in which he argues that low levels of inequality imply a big demand for medium skilled intensive goods, "providing a bridge over which low skilled dynasties may transition to the high skilled sector in the long run". Low-income families can break social mobility barriers, what is known as intergenerational mobility, and move from poverty to prosperity. What does the model describe? An increase in the wages of those workers manufacturing goods will permit parents' investment on education for their children. On the other hand, under high inequality, "the initial lack of demand for medium skilled

labour breaches this bridge from poverty to prosperity and inequality persists", these inequalities are an impediment for the accumulation of human capital. Can this model be applied to the distributions of land in Spain? If we compare the initial differences in the distribution of the lands in the north with 'minifundios' (egalitarian distribution of land) and in the south with 'latifundios' (big extensions of land with low class workers), we can explain through the inequality hypothesis differences in literacy rates in the beginning of the 20th century (for example differences in land distribution in Cataluña and Andalucia). There is a quote by the World Bank, previously cited by William Easterly[xv] where they say "we now have considerable evidence that equity is also instrumental to the pursuit of long-term prosperity in aggregate terms for society as a whole". At the beginning of the 20th century, nearly 75%[xvi] of the population in Andalucia where "jornaleros" and "campesinos", enormous differences between them and the landlord existed. This polarization and big inequality would explain, for example, why Andalucia is in such a big disadvantage to those regions in the north, in comparison to the neighbouring region Murcia, which had an intermediate distribution of land between 'minifundios' and 'latifundios', and as we have seen in Table 4, at a 10% significance level there are no statistically significant differences between the test scores in Murcia and Madrid.

Acemoglu, Johnson and Robinson (2002)[iv], disserted about the creation of good institutions and bad institutions in North and South America respectively, and the consequent development and prosperity of the different regions. Literacy rates in 1900 can be interpreted as a proxy of social institutions at the beginning of the 20th century. This would explain part of the persistence of the differences between the regions in the north and in the south. There might also be a cultural component intrinsic to each individual and to the environment in the region. Culture might affect individual preferences and incentives, providing another path in order to explain the correlation between literacy rates in 1900 and PISA results 2009.

Summing up what has been concluded in block 3, the correlation between PISA results 2009 and literacy rates in 1900 is very interesting. LIT_RATE_1900 is statistically significant at a 1% significance level. A discussion has been made in order to try to figure out the different possibilities for the persistence of differences across regions and how they were triggered; why differences in distribution of land have an impact on long run accumulation of human capital, why some regions developed earlier

and seem to be ahead at least in education output terms; the possibility of development of good and worse institutions or maybe the existence of a cultural component which maintains the differences across regions.

There is a hope for optimism, differences have been slightly converging since the year 2000 and will continue to do so with the increase in the levels of parents' education, with policies oriented to the improvement in education outputs, and maybe in a few years there will be no statistically significant evidence for the differences across regions.

CONCLUSION:

If all the regions in Spain had the same family background, similar individual characteristics and similar household environment, would the differences in education outputs change?

Results published by PISA show big differences across regions in Spain. As we have seen in the human capital models built, individual characteristics and family background play an important role in determining students' scores. Controlling for these two factors significantly reduces the differences across regions but still some of them are statistically significant. On the one hand, sex, attendance to pre-primary school, family structure, parents' education, literary culture at home, a desk to study, … as seen in Block 1 all have a statistically significant impact on students' test scores, and they have been quantified through regressions in this project. On the other hand, there is something more that maintains some of the regions at a big distance from the rest, even after controlling for those two factors. We have seen that GDP per capita (wealth of the region) has a positive impact on students' scores, although the hypothesis could be rejected at a 5% significance level. Including literacy rates in 1900 in the regressions gave unexpected results. It has a strong significant effect at a 1% level, and different possibilities for the relationship between literacy rates at the beginning of the twentieth century and PISA results 110 years later have been proposed. Inequalities in land distribution and their impact on the accumulation of human capital could be one of the explanations, literacy rates as a proxy of social institutions in 1900 gives way to another possible explanation. A cultural component affecting preferences or incentives of individuals in different regions could be part of the explanation of these differences.

This year, 2012, there will be a new report from the OECD on PISA. International exams are a good evaluation of education levels of the different countries, a measure of the stocks of human capital and a signal for the future development of the different regions. It is not acceptable that Spain is below the OECD average in the three areas; and especially in some of the regions in our country policies should be developed to improve the education levels. There is no doubt of the importance of education, and everyone independently of their social and family background, or the region they where are born, should have equal opportunities. I will like to finish my essay with a quote by Horace Mann[xvii] in 1848, *"Education, then, beyond all other devices of human origin, is the great equalizer of the conditions of men, the balance-wheel of the social machinery"*.

APPENDIX:

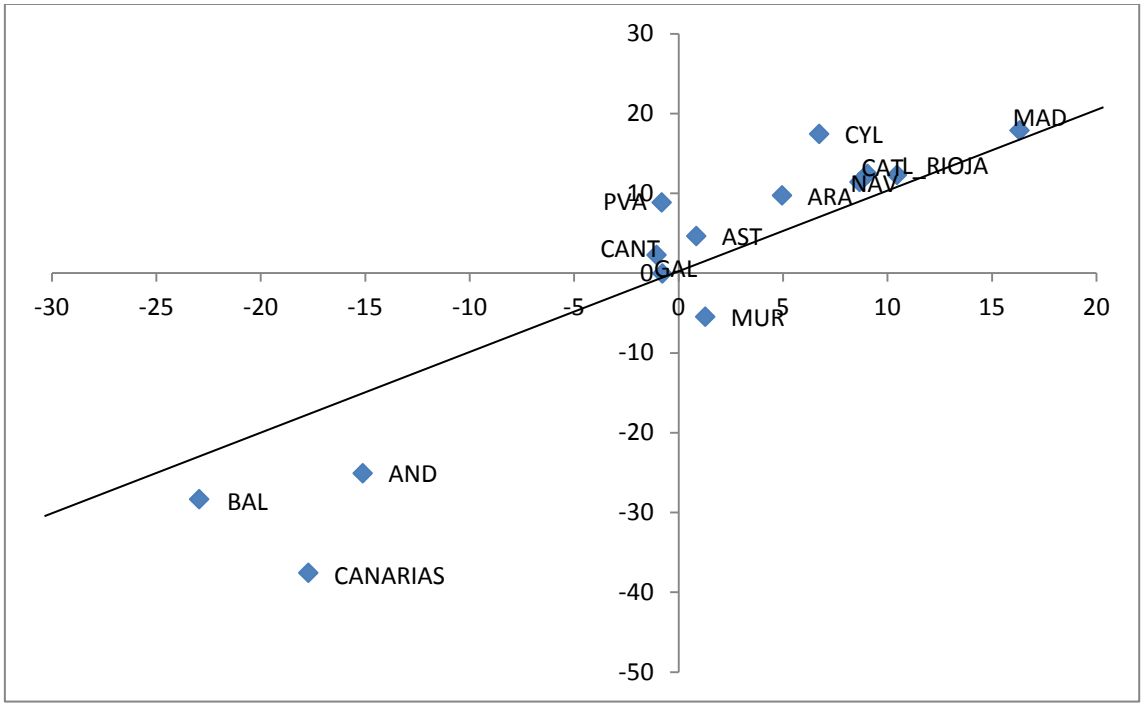

Figure 2: Conditional and unconditional differences in relation to the mean in reading

BIBLIOGRAPHY:

[i] Hanushek, E. A., and Woessmann, L (2008) "The Role of Cognitive Skills in Economic Development". Journal of Economic Literature, Volume 46, Issue 3, pp.607-668

[ii] Davis-Kean PE. (2005) "The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment". Journal of Family Psychology;19, 294–304

[iii] Haveman, R. and Wolfe, B. (1995) "The Determinants of Children Attainments: A Review of Methods and Findings," Journal of Economic Literature. 33:4, pp. 1829-78

[iv] Acemoglu, D.; Johnson, S., and Robinson, J. (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution". Quarterly Journal of Economics 117, pp. 1231-1294

[v] Engerman, S. L., y Sokoloff, K.S. (1997): "Factor endowments, Institutions, and Differential Paths of Growth among New World Economies" en HABER, S (ed.), How did Latin America Fall Behind? Stanford University Press, Stanford, CA, pp. 260-304

[vi] Ciccone, A. & Gracía-Fontes, W. (2008), "The quality of the Catalan and Spanish Education Systems: A Perspective from PISA"

[vii] PISA 2009 Results (2011), Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II)

[viii] Woessmann, L. (2000), "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence". Kiel Working Paper No. 98

[ix] Barciela López, C., Carreras Albert, Tafunell X.,BBVA (Firm). Fundación (2005). Estadísticas históricas de España: siglos XIX-XX, Volumen 1

[x] Eugenia Nuñez, C. (1992). "La Fuente de la riqueza y desarrollo económico en la España Contemporanea". Alianza Editorial, Madrid

[xi] Ray, A. and Margaret, W. (2003) PISA Programme for International Student Assessment (PISA): PISA 2003 Technical Report

[xii] OECD (2009) "PISA Data Analysis Manual: SPSS Second Edition", Paris: OECD Publishing

[xiii] Dominguez Martín, R. (2002). "Desigualdades sociales y crecimiento económico regional en España a largo Plazo", pp. 133-135 and 142-144

[xiv] Mani, Anandi, (2001). " Income Distribution and the Demand Constraint," Journal of Economic Growth, Springer, vol. 6(2),pp 107-33

[xv] Easterly, W. (2007). “Inequality Does Cause Underdevelopment,” Journal of Development Economics 84 (2), 755-776.

[xvi] Solana Ruiz, J.L. (2000) “Las clases sociales en Andalucía: un recorrido sociohistórico”, Gazeta de antropología No.16

[xvii] Mann, H. (1848). “The Republic and the School: Horace Mann on the Education of Free Men”. Report No. 12 of the Massachusetts School Board. Ed. Lawrence A. Cremin. New York (1957). 84-97