A Work Project, presented as part of the requirements for the Award of a Masters' Degree in Economics from the Nova – School of Business and Economics

Stratification and peer effects: An analysis of Lisbon public schools

Ana Rita Coelho de Azevedo
Student number 625

A project carried out under the supervision of:
Ana Balcão Reis, Maria do Carmo Seabra and Luís Catela Nunes

January 7th, 2015

Stratification and peer effects: An analysis of Lisbon public schools

Abstract

Educational stratification has been a difficult subject to deal with having yet no study shown a quantitative measure of it. Using the idea of distribution comparison a measure based on parents' education is built for the primary schools in Lisbon. Upon the confirmation that Lisbon is stratified, I use the measure of peer effects based on stratification and determine its impact on test scores, concluding that the existence of stratification improves scores of students in schools with more educated parents and decreases scores of students in schools with less educated parents. Moreover, using fixed effects I derive the conclusion that the measure of peers' characteristics helps explain most of differences among schools.

Keywords: stratification, education, stratification measure

Acknowledgments

The author would like to thank DGEEC for providing access to the data and Professors Ana Balcão Reis, Maria do Carmo Seabra, Luís Catela Nunes

1 Introduction

Education is at the basis of every modern society, playing a never-ending role in the status attainment process. As is widely recognized the social background in which individuals grow up significantly determines later socioeconomic conditions. This positive relation is argued in research on social stratification as being highly linked to educational performance (Duncan and Hodge, 1963; Eckland, 1965). Thus, "educational stratification exists as an area of interest because of the recognized importance of educational institutions in the stratification systems of societies" (Hauser, 1970).

Previous research on socioeconomic stratification has identified three primary mechanisms in which stratification of students occurs. First, stratification emerging from residential location choices that is connected to school choice in the area of residence. Second, parental choices to send children to private (or non-traditional public schools) rather than public schools and lastly, public schools resort to sorting students through tracking policies.

The literature has been keen in analyzing educational stratification by determining the impact of socioeconomic determinants on the success of students across schools. Stratification in itself, however, is a difficult subject to grass having yet no study on educational grounds tried to find a quantitative measure of it.

Basing myself on the stratification literature, more specifically Yitzhaki and Lerman (1991), I try to make the first steps into this subject by obtaining a measure of school population characteristics. This is done by distribution comparison for primary schools in Lisbon, looking for the difference between school and city population. Such was employed based upon parents' education for its vital link to children' achievements. Of this analysis, there is the confirmation in the idea that Lisbon is at some degree stratified. This method however, does not return school levels of stratification rather than just a big picture of what the city looks like and because of the way it is obtained, at its core seen as a peer effect.

Also, to infer on the effect on test scores a standard education production function model is applied arriving to the conclusion that as stratification exists, students in schools with more

educated parents tend to perform better and students in schools will less educated parents worse.

This report is structured as follows. Section 2 presents the relevant literature on the matter. Section 3 provides a review of the Portuguese primary education system. The dataset is presented in section 4, as well as a brief summary of the data. Section 5 discusses the concept of stratification while section 6 deals with the proposed measure. Section 7 details an empirical framework and its results and a conclusion is given in Section 8.

2 Literature Review

On the topic of educational stratification there have been three distinct paths addressed in the literature. First, families may self-select into different traditional public schools since housing markets encourage residential segregation which in turn results in school-level stratification. Second, some families may select out of traditional public schools in the presence of private or nontraditional public schools. Finally, traditional public schools themselves may choose to sort students through tracking.

Hoxby (2000) and Urquiola (2005), based on the Tiebout variation, document that residential choices by parents result in a nonrandom stratification of students across public schools that vary in inputs and outcomes. Moreover, on the margin, parent's choices regarding districts of residence do affect their children's peer groups but most of the segregation along race and income happens at the school, not the district level. Furthermore, "resulting increases in district homogeneity have little net effect on achievement, per-pupil spending, or productivity" (Hoxby, 2000).

In the debate of school choice, the argument is given that public schools are inefficient local monopolies, educational quality would improve if parents were allowed to freely choose between schools. Yet, while acknowledging the potential positive effects of school choice, its effects on inequality are a concern, especially for those students remaining in the public sector, resulting choice in greater segregation of students by ability and socioeconomic background.

Empirically there has been a consensus that choice might result in higher segregation.

Scholars acknowledging this have claimed that "if choice leads to a higher degree of sorting by income and ability and peer effects matter, then, the distribution of educational benefits is likely to be quite unequal" (Epple & Romano, 1998; Epple, Figlio, & Romano, 2004). Moreover, Epple, Figlio, & Romano (2004) presents evidence that confirm the results of the theoretical model developed in Epple & Romano (1998) that private schools engage in price discrimination that leads to sorting on observed ability within the private sector, with the highest income and highest ability students in the top private schools. Lankford and Wyckott (2001) also conclude, by estimating a model of public-private school choice, that students who choose to move to private schools are some what different from public school pupils left behind on academic and family dimensions.

At school level, Bifulco and Ladd, (2006) and Saporito, (2003) claim increased ethnic and racial stratification between schools as a consequence of school choice because parents tend to choose schools with racial and economic compositions similar to their own backgrounds. Saporito (2003) looks at magnet schools, finding no empirical evidence that choice policies reduce segregation by race and class, even arguing the opposite. He also postulates that the "choices of white and wealthier students lead to increased racial and economic segregation in the neighborhood schools that these students leave". For charter schools in North Carolina, Bifulco and Ladd (2006) conclude that sorting of students based on racial and socioeconomic background contributed to the poor performance of these schools.

Research on countries where the national voucher systems operates, such as Chile and Sweden, also arrive to the conclusion that choice increases stratification. Böhlmark and Lindahl (2007) analyze the educational reform occurred in Sweden in 1992, where the country transited from a assigning students to their closest public school to a system allowing free choice between public and private schools. They concluded that, the competitive forces of the reform induced higher achievement but also higher segregation for migrant students as parents with higher levels of education tend to choose private schools for their children. Likewise, in Chile, it is argued that the voucher program led to increase sorting, as the best public school students left for private schools (cream-skimming). (Hsieh and Urquiola, 2006; McEwan, Urquiola and

Vegas, 2008; Mizala and Torche, 2010;) Moreover, Hsieh and Urquiola (2006) state that the shift to the private sector did not produce achievement gains.

In a different spectrum, public schools frequently attempt to provide differentiated products within individual schools whether trough partial or complete tracking. In most educational systems students are placed in distinct tracks or curricular programs. At the secondary level, the most common distinction is between the academic tracks that prepare students for higher education, and those tracks that prepare them for immediate entry into the labor force. Track placement is determined largely by the students' prior achievements. But because student achievements are correlated with their socioeconomic origins, especially with parent's educational attainment, students from less privileged strata are more likely to attend non-academic tracks (Heyns, 1974; Alexander, Cook and McDill, 1978; Rees, Argys and Brewert, 1996). Track placement, in turn, affects their subsequent educational attainment and enlarges inequality between social strata in subsequent attainments.

The issue of ability tracking was most directly approached by Epple, Newlon and Romano (2002), whose model exogenously set levels of tracking within public schools and demonstrates that tracking serves to retain relatively high-income, higher-ability students who qualify for the higher track, meaning that tracking on the part of public schools results in a more intense competition for higher-ability students.

3 Portuguese Primary Education

The choice to focus on primary education stems for it being the basis of the educational system, where every individual is introduced into the basic studies. This analysis of stratification right at the earliest stage of education seemed the right approach since its possible effects on every other stage stage given the cumulative power of education. Also, in primary schools the same curriculum is offered to all students, existing no tracking.

In Portugal, basic education encompasses nine years of schooling, being compulsory and divided into three cycles. It follows a National Curriculum whose standards are defined in the form of general and specific competences. The first cycle of basic education covers years 1st

to 4th, the second cycle years 5th to 6th and the third cycle years 7th to 9th. For the purpose of this work the focus will be on the 1st cycle of basic education which is taught in public or private primary schools.

The enrollment in the 1st year of primary education is compulsory for children who are six years old by 15 September. However, guardians of children who reach six between 16 September and December 31, may submit an application to the school governing body to be allowed for registration in that school year, being the latter conditional upon vacancy. Moreover, all students who attend primary schools have to take national exams at the end of the cycle (4th grade), not having, in the time period in analysis, a weight on students final grade[1]. The grading system goes from 1 to 5, being the threshold to "pass" the exam equal to 3.

4 Data

The dataset used in this empirical study includes information about students attending primary public and private schools in Portugal from 2007 to 2012. The set of variables allows us to know what school a given student is in each year, his grade, as well as personal information about themselves and their parents. Moreover, it is possible to combine this with achievements, using the JNE (Júri Nacional de Exames) database, providing test scores on national exams that occur at the end of 4th grade.

The choice to focus on public schools comes, first of all from the richness of the data, that having information for private schools is much more extensive for the public sector. Also, public schools are free of charge (no tuitions), so I look at stratification that is the result of other mechanisms different from price. Apart from this, the data was restricted to the municipality of Lisbon. It makes only sense to test stratification within a delimited region where children may be allocated to different schools in different ways and Lisbon seemed an excellent case to start with for its great diversity steaming from being the capital city, having a large number of public primary schools; around 90, which is highly important given that

[1] From 2013 on such exams will have a weight on the student final grade

stratification is easier to occur, and therefore test, as the number of schools increases because the easier it is for pupils to be aggregated with similar peers.

Having this into account, there are 92 680 students in 91 primary schools in Lisbon during the time period in analysis. Moreover, the JNE database includes 20 104 observations for 4th grade students in Lisbon. A table with summary statistics can be seen below. The percentages of females among primary and 4th grade students is around 48%, while immigrant students account for 7%, while the percentage of subsidized-lunch students is 40%. Also, average test scores for Portuguese and Mathematics are quite similar and around 3, while the percentage of failed students is much higher for Mathematics.

Table 1: Summary statistics for the 2007-2012 period

	4th grade students	Primary students
Student Characteristics		
Age (in years)	9.37	7.84
Female (%)	47.83	47.45
Immigrant (%)	7.26	7.03
Subsidized-lunch (%)	39.84	39.83
Family Background		
Immigrant parents (%) (a)	16.84	17.87
Parents' highest level of education (%) (b)		
Primary or less	27.04	37.08
2nd cycle of basic school	12.17	13.57
3rd cycle of basic school	21.75	15.78
Secondary education	21.52	14.74
Higher education	17.52	18.83
Unemployed parents (a)	21.32	22.34
Test scores		
Test scores math	3.19	-
Test scores portuguese	3.15	-
Fail math (%)	13.34	-
Fail portuguese (%)	6.1	-
No. Observations	20104	92680

Note: This reports summary statistics (mean) for the 2007-2012 period for age and test score results. Other variables are expressed as a percentage. The units are explained in front of the variables. (a) at least one parent; (b) average of both parents

5 The concept of stratification - descriptive data

In its essence, stratification is "the division of a society into a number of strata, hierarchically arranged groupings" and "groups form well-defined strata to the extent that their members differ from the rest of the population" [2]. Before entering in the measurement of these effects one needs to start from the basics and see how students are distributed across schools in Lisbon.

To do so, a characterization of schools based upon key socioeconomic variables of interest such as parents' education, nationality and subsidized-lunch[3] is foremost important. The point being to observe the variability across schools because from the notion of stratification presented above, if schools are essentially the same in terms of students' characteristics there is no point to talk about stratification because there would be none: pupils in each school would be the same as the rest of the population.

In light of this, for each school I looked at the % of students whose parents have at most the 9th grade; the % of students with parents with at least an undergraduate degree; the % of students with non-Portuguese parents and lastly the % of students with subsidized-lunch [4].

A first glance at these variables provide the conclusion that there is, in fact, variability between schools, as can be seen in the graphs below. For instance in 2012 [5], some schools bear 80% of students whose parents have at most the 9th grade, being such schools associated with poorer/more problematic areas of Lisbon. On the other side of the coin, schools that have students with more educated parents are located in better areas of Lisbon and few are the schools in which 40% of students whose parents have an undergraduate degree. For kids of immigrant parents, most schools encompass less than 20% of them, being, nevertheless, the higher percentages in this measure related with schools where parents are less educated and consistent with studies that show that the area of the historic center of Lisbon is the location

[2] Yitzhaki and Lerman (1991)

[3] Called, SASE (Serviços de Acção Social) in Portugal. It is divided into 3 levels, A, B and no-sase

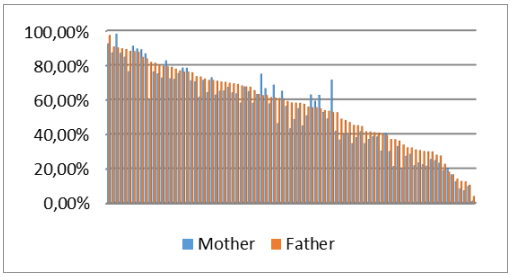
[4] From this, it is essential to exclude schools in which observations are not available or unknown. This requires losing information, being the analysis susceptible to the variable of interest in use. In table A1 from the appendix it is possible to check how much information is lost when characterizing schools for each of the measures mentioned above.

[5] The result are quite similar to other years in terms of direction of the analysis done for 2012.

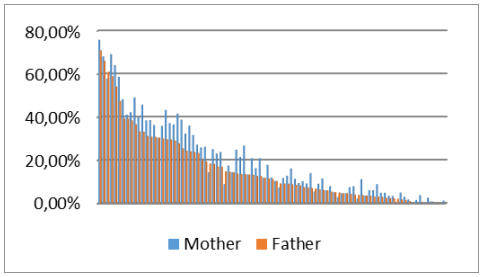
with stronger concentration of foreign populations[6] .

Figure 1: **Characterization of schools**

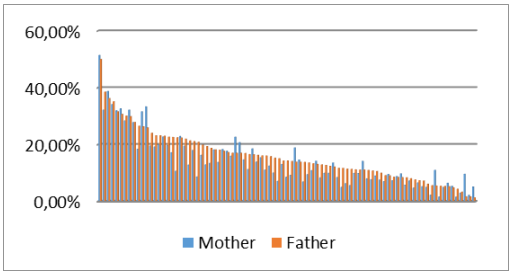
(a) **% of students per school with parents with at most the 9th grade, in 2012**



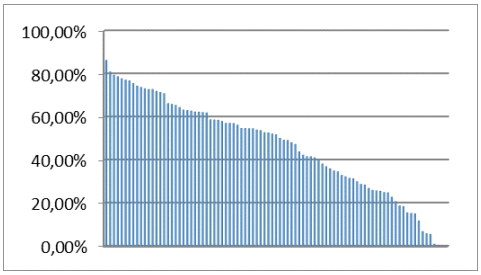
(b) **% of students per school with parent with at least an undergraduate degree, in 2012**



(c) **% of students per school with non-Portuguese parents, in 2012**



(d) **% of students per school with subsidized-lunch, in 2012**



The striking idea given by these statistics is that there is scope to analyze educational stratification in Lisbon schools. As a starting point on measuring stratification I decided to focus the analysis onto one key characteristic which according to the literature is highly influential on pupils' achievements: parents' education (Hanushek, 1986).

The most straightforward explanation for the connection between parents' education and their children's academic achievement is that parents absorb something during schooling that ends up influencing their values and knowledge impacting the the ways of interaction with kids about school. Also, more educated parents usually have higher expectations for their children's education, thus predicting greater educational attainment for these kids (Alexander et al., 1994).

Moreover, education influences family choices, the types of jobs both parents are likely

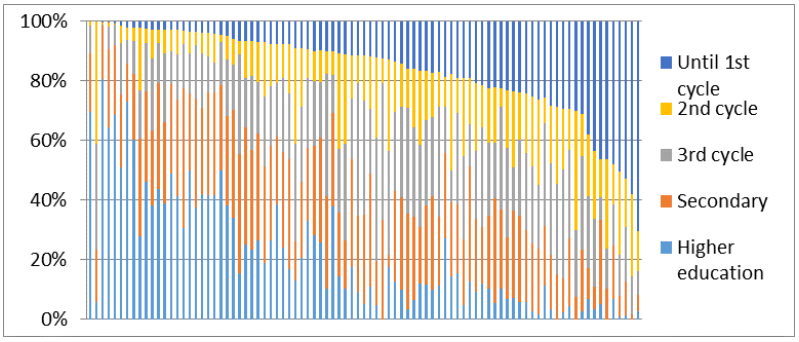
[6]Hortas, Martins and Dias (2014)

to have and, thus income and living opportunities, which impacts the types of school and neighborhood that their children will be exposed to (Coleman, 1987), that directly influence children's educational achievement.

To properly analyze the distribution of students according to parents' education I aggregated the several levels of education, combining higher, secondary, and several steps of basic education (without qualifications and 1st, 2nd and 3rd cycles).

The stacked graph below from 2012 (being quite similar for all years) shows the percentage of students, in each school, with fathers[7] with a given level of education, it has been sorted by the lowest level of education, in a way that the schools more to the right have more students with less educated parents. It seems to exist a polarization at the highest and lowest level of education: there are at one end a few schools with really well educated parents and at the other schools with poorly educated parents. Nevertheless, this shows both huge variability and that there is no perfect stratification: the large majority of schools have parents with all levels of education.

Figure 2: **% of students by school with fathers at a certain level of education, in 2012**



6 Measuring Stratification

The first prominent thought in deriving a measure that allows looking at stratification streams from comparing distributions following the reasoning of Yitzhaki and Lerman (1991). Thus,

[7] similarly for mothers

it should try to provide answers for the following question: How much does the distribution of parents' education in a school diverge from the municipality distribution? Putting in another way, students distributed in a way that their peers are much like their own (in terms of parents' education) but differing from the general population? Moreover, in this analysis the database was restricted to students for which information on parents' education was available (accounting for 67% of the existing information for Lisbon).

To address this, one should first look at the position (rank) of students within their own school and the students' position in the overall population, in this case among the total primary students of Lisbon.

Let y_{ij} be the combined years of parents' education[8] of student i in school j, with n members. The rank (r) of the parents' education (ranked from less to more educated) will give us the position of student i among all students in a given school, for example, if student i is the one with less educated parents he will be ranked one in his school while the best will be the nth student. It is, nevertheless, vital to look at a percent rank to be able to compare both results the student position in the school and in the whole population, using the formula, $PR = {}^{(r-1)}/_{(n-1)}$, in order to yield for the unique middle the value of 0.5 when the sample size is odd and treating tails symmetrically. This approach, returns values between 0 and 1, where zero reports the worst ranked student (with less educated parents). Such was employed both at school (PR_{ij}) and population level (PR_{ip}). Moreover, for a given student, the key is, thus, to look at the difference between both ranks:

$$F_i = PR_{ij} - PR_{ip} \tag{1}$$

Also, to clearly examine how much the distribution of a given school diverges from the municipality the final measure should account for how all students of a school perform within the school versus their performance within the city. To do so, one should look at the sum of all differences between these ranks and normalize it by taking into account school size.

[8] An average of both parents' education, provided that was information available for at least one of the parents.

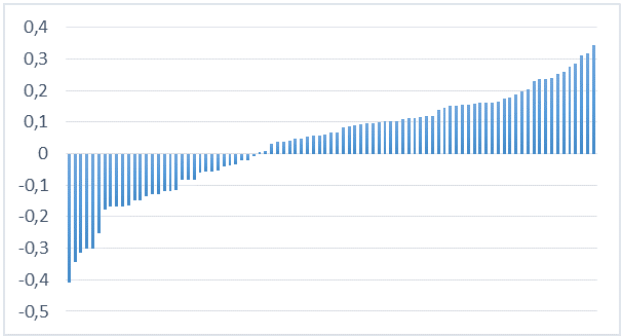
$$S_j = \frac{\sum F_i}{n} \tag{2}$$

, where the sum is over all students in school j.

From here it is to expect that, for schools that have students with less educated parents, the school rank of students is higher than the overall one, returning positive values of F_i. Antagonistically, for schools that that have students with more educated parents, the overall rank always surpasses school ranks, being F_i negative. To better understand this, take two schools, the very "best" in the city, with more educated parent and the very "worst", meaning schools located at the extremes of the city distribution. The "average" ranked student in the "worst" school would be much closer to the bottom in the whole population (school rank>city rank), while the "average" ranked student in the "best" is much closer to the top overall (city rank>school rank). Furthermore, S_j returns values between -1 and 1.

It would be expected that in the absence of stratification in the city, all schools would achieve values near zero, having no differences between the ranks of students in the school and their rank in the overall population, essentially all students in the city would be similar. there seems to be evidence confirming the idea that Lisbon is stratified. There is higher segregation at the extremes, with a gathering of a few schools with most educated parents and schools with the least educated, while the other schools in the city appear to have a more or less uniform pattern between them, even with perfect stratification schools in the middle of the distribution would have S_j near 0.

Figure 3: S_j**by school**



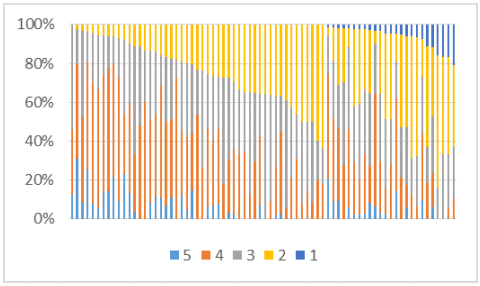
It is to mention, however, that S_j does not provide a per school value of stratification. It is only possible to infer on the stratification of the city: taken together, the values for all schools are an indicator proving information that there is some kind of stratification in the Lisbon because they differ from zero, but we do not know its size or if a given school is more stratified than another. The measure may be understood as a measure of peers' characteristics, because the essence of its calculation: it is because stratification exists that students are aggregated into schools as they are.

Another and simpler way of thinking about stratification would be just to look at the average education of parents per school. Nevertheless, it does not allow to take into consideration the heterogeneity within schools as comparing the distribution between school and city does, for the fact that we can achieve the same average results with really homogeneous schools or with schools that, being quite heterogeneous, bear the loads of the extreme values.

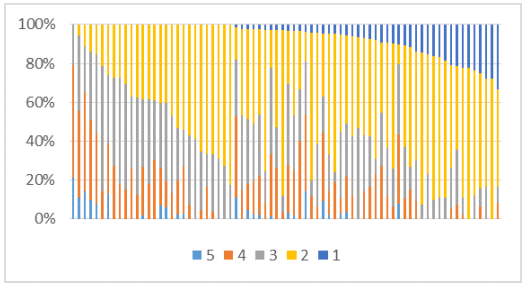
It is also interesting to look for stratification in the outputs rather than just the inputs of educational achievement, and check if there is a matching pattern or not between the two. To do so, the process implemented with parents' education was repeated for test scores. The graphs below show the characterization of schools according to the grading levels (sorted by the negative grades), showing also a great variability across schools.

Figure 4: **% of students by school with each of the test score levels**

(a) **For Portuguese test**



(b) **For Mathematics test**



Adding information on test scores conditional on the measure of peer effects based on stratification (S_j) shown below, one concludes that schools with less educated parents seem to perform worse (S_j>0), as the percentage of negative test scores in mathematics is 18% (16%

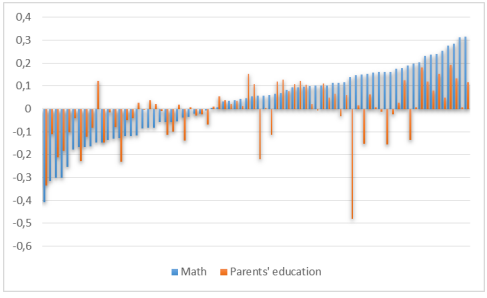
for Portuguese) higher when comparing to schools with more educated parents (S_j<0). Also, the highest score is significantly lower by 11% (6.7%).

Table 2: Highest and lowest test scores percentages in Portuguese and Mathematics conditional on the peer measure

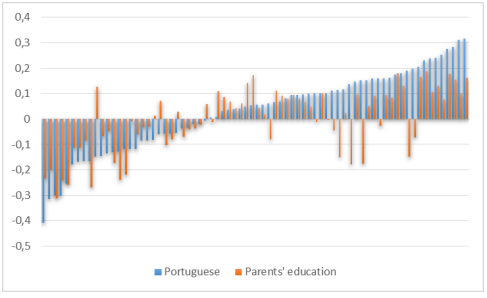
		Math	Pt
S_j>0	% of negative scores (1 and 2)	32.49	24.27
	% of highest score (5)	6.49	3.2
S_j<0	% of negative scores (1 and 2)	14.43	8.26
	% of highest score (5)	17.91	9.9

Apart from this, to see if Lisbon is also stratified by achievements and to compare it to the pattern of stratification found before (for parents' education), the same logic in measurement was employed but using now as input test scores instead of parents' education. This means that it accounts for how much the distribution of test scores in a given school diverges from the distribution of the city. The graphs below show the combination of both results being sorted by the one using parents' education.

Figure 5: **Stratification measure by parents education and test scores**



(a) **Comparison between stratification measure by parents education and math test score sorted by the parents education one**



(b) **Comparison between stratification measure by parents education and portuguese test score sorted by the parents education one**

By looking at them one can see that it also seems to exist some sort stratification by achievement being the pattern not followed by the one found for parents' education. Nevertheless these results, provide the conclusion that students are segregated in Lisbon by their parents' education but if the variation in achievement is a consequence of this (which we don't know) than there is not necessarily a segregation by achievement. This is just the outcome. But how does stratification influences test scores? For this, one will apply the empirical framework discussed below.

7 Empirical Framework and Results

To examine the effect of stratification on test scores a standard education production function model is applied. This estimate test scores, as a function of the cumulative impact of school inputs, the peers of the student, their individual characteristics and family. The model to be estimated is:

$$T_{ijkt} = X_{ijt}\beta + \gamma D_{jt} + \delta S_{jt} + \varepsilon_{ijkt} \tag{3}$$

where T_{ijkt} is the test score for student i at school j for subject k in year t, X_{ijt} is a vector of individual and family background influences, D_{jt} is vector of school inputs, S_{jt} is the measure of peer characteristics based on stratification as mentioned before and, ε_{ijkt} is an error term.

The baseline regression was estimated with results showed further below. The dependent variable is either the test score from 1 to 5 on national exam in Mathematics or Portuguese. Age refers to students' age until September of the school year. Gender is a dummy variable equal to 1 if the student is female and zero otherwise.

Following the reasoning of Chiswick and DebBurman (2003) that immigrant generation are significantly important for educational achievement, the variable Immigrant was generated, being equal to 0 if the student is Portuguese of Portuguese parents (native), 1 (second generation immigrant) if the student was born in Portugal from foreign-born parents and 2

(first generation immigrant) if the student is foreign born from at least one foreign-born parent.

Subsidized-lunch is divided into three categories, 0 referencing to students with no help, 1 for intermediate students and 2 for students with the highest level of subsidy. Parents' employment situation varies from 1 – Employed, 2 – Unemployed, 3 – Student, 4 – Retired and 5 – Other. Parents' education is the combining years of education (average) of both parents for each student and lastly, S_{jt} is the measure of peer characteristics . There were also included as controls dummy variables for years.

Table 3: Estimates of empirical model

	Math	Pt		Math	Pt
Age	-0.16976**	-0.20526**	Employment Mother		
	(0.01169)	(0,0102)	2	-0.04685	-0.0395**
Gender	-0.12172**	0.178128**		(0.0208)	(0.0181)
	(0.14632)	(0.0127)	3	-0.29454	-0.10579
Immigrant				(0.1332)	(0.1159)
1	-0.12837**	-0.01018**	4	-0.04719	0.039204
	(0.02878)	(0.0251)		(0.1659)	(0.1445)
2	-0.10387**	-0.06777**	5	0.029703	0.076863**
	(0.03357)	(0.0292)		(0.0313)	(0.0272)
Subsidized-lunch			Parents' educ	0.031299**	0.023155**
1	-0.10554**	-0.08281**		(0.0011)	(0.0009)
	(0.0241)	(0.0211)	S_j	-0.69503**	-0.68276**
2	-0.16333**	-0.15775**		(0.0528)	(0.0460)
	(0.0212)	(0.0184)	Year dummies	YES	YES
Employment Father			R^2	0.2907	0.2468
2	-0.04736	-0.04456*	Adj-R^2	0.2896	0.2456
	(0.0292)	(0.0254)	No. Obs.	13347	13347
3	0.149064	0.288623			
	(0.2061)	(0.1795)			
4	0.018256	0.001415			
	(0.0771)	(0.0671)			
5	-0.04822*	-0.00916			
	(0.0277)	(0.0241)			

Note: standard errors are reported in parentheses. ** - significant at 5%, * - significant at 10%

As shown by the results above for Mathematics and Portuguese test scores students' age, gender, subsidized-lunch allocation, parents' education, immigration and the stratification measure are highly significant. Students that have subsidized-lunch tend to underperform while comparing to students that don't. Such result is expected, as students with such aid

are from poorer or more troubled families, factors that are recognized to negatively influence children' achievement.

Moreover, as age increases by 1 year, the test score is expected to decrease by 0.2, ceteris paribus. Age is also capturing students that are repeating school years, thus making sense such result. Also, being a girl hurts test scores in Mathematics but significantly increases scores in Portuguese. Being first or second generation immigrant also hurts changes of better grade while compare it to native students but surprisingly enough the effect is lower for Portuguese. The striking and crucial result is that students in schools that diverge from the city distribution negatively, tend to perform worse: the fact that there is stratification improves test scores of students in the schools with more educated parents and decreases scores of those in schools with less educated parents.

Furthermore, is also important to realize that including both the parent's education for each student and the measure of stratification that is based upon this variable returns results of significance for both of them, proving that in their essence they are measuring different impacts.

As mentioned before, another way of looking at stratification would be just to apply an average of parents' education by school. In reality, this is highly correlated (-0.97) with my measure, so they both capture the same explanatory effect, however, also as stated before, averages do not allow heterogeneity among schools to be "felt", and moreover, when including both variables and parents' education on the baseline regression it is the average education that loses significance, which empowers the measure of peers based on stratification as having more explanatory power in determining achievement [9].

At this point one already tackled the most important issue of this work: Lisbon is stratified and the measure of characteristics of the school population based on stratification significantly influence test scores. Nevertheless, the explanatory power of the model showed above (29% of the variation in mathematics test scores are explained by the variation in the model, 24% for Portuguese) makes us question what else is beyond this. What more explains the variability of test scores? Is it more prominent within or between schools?

[9] To see the result of the model with both variables please check Table A2 in the appendix.

First of all, how much of test scores is explained by Sj itself? Running a simple regression one discovers that 17.97% of the variation in mathematics test scores (13.06% for Portuguese) is explained by the variation in this peer measure. This number is quite significant and more relevant for Mathematics than Portuguese, which may be because schools are relevant and peers matter more given hardness of the subject. But being this something schools cannot control, how much do schools really matter to explain achievement? Do they add more than this? Many authors have concluded that schools matter but not as much as other factors for resins that are not yet clear.

Since school effects are potentially unobservable they should, therefore, not rely solely on observable characteristics. For this reason, a model with school fixed effects deals better with such situation. The fixed effects φ_j to be a group specific constant term in the regression equation:

$$T_{ijkt} = \varphi_j + X_{ijt}\beta + \varepsilon_{ijkt} \tag{4}$$

The term φ_j is presumed to capture the unobservable that differentiate school units. Basically, this implies that all differences between schools are fixed over time and represented as parametric shifts of the regression function. Thus, every variable that is school specific cannot enter the model. Since we are not dealing with panel data, this can be achieved by adding school dummy variables.

$$T_{ijkt} = X_{ijt}\beta + \varphi_1 D_1 + \varphi_2 D_2 + ... + \varphi_n D_n + \varepsilon_{ijkt} \tag{5}$$

Where D_{i1} is 1 when j=1 and 0 otherwise, D_{i2} is 1 when j=2 and 0 otherwise, and so on, being just dummy variables indicating the groups in the fixed effects model above.

Applying this logic, but using only school fixed effects and no other explanatory variables[10] , 25.48% of the variability in Math (19.51% for Portuguese) test scores is explained by the variability in school effects. This leads us to conclude that more than 75% of test scores variation is explained, not between schools but within each school. This appears to go in

[10]There is nevertheless, the need to use year dummies as controls in all of these steps.

line with the literature in the matter of school effects where Kramarz, Machin and Ouazad (2009) concluded "pupil heterogeneity to be a more important determinant of achievement than school quality" even though both inputs were statistically significant.

By comparing results of both models, where the school variable is regarded as being either the measure of peer effects (Sj) or the fixed effects we take a lesson that schools encompasses more than just peer effects. Looking at the R-squared reported above for the simplest specification possible, schools add around 7.5% (for Math the R-squared goes from 17.97 to 25.48). To further stress this point, some specifications were done, taking as control parents' education. From the table below, reporting the R and adjusted R-squared we can also infer on the significant effect of parents' education on kids' achievement. Including this into de models helps explain more 6 to 8% of the variability in test scores (for Math, for instance, in the model with peer effects, R-squared moves from 17.97 to 26.66, while for fixed effects it changes to 31.25 from the previous 25.48).

Table 4: R^2 and adjusted R^2 of specification models

		Mathematics test score		Portuguese test score	
	Parents' education	R^2	Adj-R^2	R^2	Adj-R^2
S_j	-	17.97	17.94	13.06	13.03
S_j	Yes	26.66	26.62	20.01	19.97
School fixed effects	-	25.48	25.12	19.51	19.13
School fixed effects	Yes	31.25	30.78	22.96	22.44
School fixed effects	Yes +subsidized-lunch, age and gender	33.17	32.69	27.23	26.71

Note: all tests have as controls year dummies

If we further develop the fixed effect model adding variables found highly significant before such as subsidized-lunch, age and gender the variability of test scores in math is now explained by 33% of the model variability, a gain of around 2%.This final specification of the fixed effects model is reported below where one can see that all variables are highly significant and follow the directional effect in explaining achievement found before.

Table 5: Estimates of fixed effects model

	Math	Pt
Age	-0.16917**	-0.20392**
	(0.0012)	(0.0098)
Gender	-0.11757**	0.184207**
	(0.0141)	(0.1242)
Subsidized-lunch		
1	-0.11242**	-0.08558**
	(0.0234)	(0.0206)
2	-0.,16881**	-0.15063**
	(0.0202)	(0.0178)
Parents' educ	0.031334**	0.022593**
	(0.001)	(0.0009)
Year dummies	YES	YES
R^2	0.3317	0.2723
Adj-R^2	0.3269	0.2671
No. Obs.	13760	13760

Note: standard errors are reported in parentheses. ** - significant at 5%, * - significant at 10%

To conclude, even though stratification exists in Lisbon and the peer measure that streams from it significantly determines how students perform there is more to the story than that. School effects do not only depend on this. Nevertheless, peers explains much of the differences between schools (the explanatory power provided only by peer effects accounts for 70% of the one arrived for school fixed effects). Furthermore, there is clearly more variation within schools than across them, which is in line with the literature findings that pupil heterogeneity matter more, that may be the reason why the pattern of stratification of inputs and outputs does not match.

8 Conclusion

For the purpose of assessing stratification in primary schools in Lisbon, I developed a measure based on parents' education. This measure is grounded on distribution comparison between school and city population. It provides something that can be interpreted as a measure of peers' characteristics and its per school values provide an indicator that there is some degree

of stratification in Lisbon.

Further, to infer on the impact of stratification in student achievement a standard education production function model was applied controlling for students characteristics, socioeconomic background and the measure of peer characteristics. The estimates were significantly clear on the negative impact of the measure of peer characteristics (that depends upon the existence of stratification) on test scores for both Portuguese and Mathematics. To further stress school effects and see if test score variability was higher among or between schools, a model of fixed effects was performed deriving the result that 75% of test scores variation is explained not between but within schools and that stratification itself helps explain the majority of the variability across schools. This is a striking result, since it helps detangle part of what can been seen as school quality into something schools themselves cannot really control and be of help to policy makers. Also, the measure of peers' characteristics has more explanatory power than the simple average of parents' education.

In a nutshell, there is stratification and the fact that it exists negatively affects students test scores in the schools where students' parents are less educated, helping explain the majority of the variability across schools.

The relevant contribution of this paper, even with all its simplicity, is in it being one of the first to approach educational stratification by trying to quantify it, being a start for future research. In the future, it would be relevant to use other datasets in order to check the robustness of the measure, applying it using other socioeconomic variables or a combination of them. Further it should also be applied, in the case of Portugal, to other municipalities to see how regional differences play a role in educational stratification. Moreover, as it cannot be stated that stratification decreases (or increases) students' per se, changing the stratification degree would affect students differently depending on the schools they are enrolled in and on their characteristics, which would also be something interesting to look at.

References

[1] Alenxander, K.L., Entwisle, D.R. and Bedinger, S.D. 1994. “When expectations work: race and socioeconomic differences in school performance, Social Psychology Quarterly, 57(4), 283-299

[2] Alexander, K. L., Cook, M. and McDill, E. L. 1978. "Curriculum Tracking and Educational Stratification: Some Further Evidence", American Sociological Review Vol. 43, No. 1, pp. 47-66

[3] Bifulco, R., & Ladd, H. F. 2006. The impacts of charter schools on student achievement: Evidence from North Carolina. Education Finance and Policy, 1(1), 50-90

[4] Böhlmark, A., & Lindahl, M. 2007. “The impact of school choice on pupil achievement, segregation and costs: Swedish evidence”. (Discussion Paper No. 2786)

[5] Chiswick, Barry R., & Noyna DebBurman. 2004. “Educational Attainment: Analysis by Immigrant Generation,” 23 Economics of Education Rev. 361–79.

[6] Coleman, J.S. 1987. “Families and Schools”, Educational Researcher, 16, 32-38

[7] Duncan, Otis Dudley and Robert W. Hodge 1963 "Education and occupational mobility: A regression analysis." American Journal of Sociology 68:629-644.

[8] Eckland, Bruce K.1965 "Academic ability, higher education, and occupational mobility." American Sociological Review 30: 735-746.

[9] Epple, D., & Romano, R. 1998. "Competition between private and public schools, vouchers, and peer-group effects" The American Economic Review, 88(1), 33 - 62.

[10] Epple, D., Figlio, D. and Roamno, R. 2004. “Competition between private and public schools: Testing stratification and pricing predictions”, NBER Working Paper Series, No.7956, Cambridge, Massachusetts, USA

[11] Epple, D., Newlon, E., Romano, R. 2002. “Ability tracking, school competition, and the distribution of educational benefits”. Journal of Public Economics 83, 1–48.

[12] Gamoran, A. 1987. "The stratification of high school learning opportunities", Sociology of Education, 60, 135-155.

[13] Hanushek, E. A. 1986. “The economics of schooling: Production and efficiency in public schools”, Journal of Economic Literature XXIV, 1141{1177. Hauser, R. 1970. "Educational Stratification in the United States." Sociological Inquiry 40:102-129

[14] Heyns, B. 1974. "Social Selection and Stratification Within Schools" American Journal of Sociology Vol. 79, No. 6, pp. 1434-1451

[15] Hortas, M.; Martins, C. and Dias, A. 2014. “Escola, Comunidade E Território: Dinâmicas Educativas Locais Na Integração De Populações Imigrantes Na Área Metropolitana De Lisboa”, INTERACÇÕES NO. 29, PP. 8-36.

[16] Hoxby, C. M. 2000. “Does competition among public schools benefit students and taxpayers?” American Economic Review, Vol. 90, No. 5, pp.1209-38

[17] Hsieh, C.T. and Urquiola, M. 2006. "The effects of generalized school choice on achievement and stratification: Evidence from Chile's school voucher program", with Chang-Tai Hsieh, Journal of Public Economics

[18] Kramarz, F., Machin, S., Ouazad, A., 2009. What Makes a Test Score? The Respective Contributions of Pupils, Schools and Peers in Achievement in English Primary Education. CEE Discussion Paper 102.

[19] Lankford, H. and Wyckoff, J., 2001. "Who Would Be Left Behind by Enhanced Private School Choice?" Journal of Urban Economics, Volume 50, Issue 2, Pages 288–312

[20] McEwan, P., M. Urquiola, and E. Vegas. 2008. “School choice, stratification, and information on school performance: Lessons from Chile,” Economia, 8(2), 1–27.

[21] Mizala, A., & Torche, F. 2010. “Bringing the schools back in: The stratification of educational achievement in the Chilean voucher system”. International Journal of Educational Development, 53, 132-144

[22] Nunes, Luís Catela, Portela, Miguel and Reis, Hugo. 2014. “Manual Da Base De Dados MISI Anonimizada” Documento de Trabalho

[23] Nunes, Luís Catela; Pereira, Diogo; Reis, Ana Balcão; Rodrigues, Guilherme and Seabra, Maria do Carmo. 2014. “Manual De Ligação Entre Registos De Alunos Na MISI - Período 2006/07–2011/12”. Documento de Trabalho

[24] Rees, Daniel I. & Argys, Laura M. & Brewer, Dominic J., 1996. "Tracking in the United States: Descriptive statistics from NELS," Economics of Education Review, Elsevier, vol. 15(1), pages 83-89, February.

[25] Saporito, S. 2003. “Private choices, public consequences: Magnet school choice and segregation by race and poverty”. Social Problems, 50(2), 181-203.

[26] Urquiola, M. 2005. “Does school choice lead to sorting? Evidence from Tiebout variation,” I 95 (4): 1310 – 1326

[27] Yitzhaki, Sh. & Lerman, R. I. 1991. ‘Income stratification and income inequality’, Review of Income and Wealth, vol. 37, no. 3, pp. 313–29.

Appendix

Table A1 - No of possible observations for each variable and in which the characterization of schools is based upon

	1st cycle in Lisbon		Parents' education		Nationality		Subsidized-lunch	
	#students	#schools	#students	#schools	#students	#schools	#students	#schools
2007	15100	82	9848	79	15050	82	15100	82
2008	15564	81	10172	81	15425	81	15459	81
2009	15760	84	10498	82	15663	83	15197	84
2010	15588	85	10641	83	15516	85	15588	85
2011	15339	89	10485	84	15296	88	15339	89
2012	15329	89	10796	89	15242	89	15329	89
	92680		62440		92192		92012	

Table A2 - Estimates of empirical model including average of parents education

	Math	Pt		Math	Pt
Age	-0.1759**	-0.21186**	Employment Mother		
	(0.01167)	(0,01016)	2	-0.07173**	-0.0622**
Gender	-0.12145**	0.17844**		(0.0205)	(0.0178)
	(0.14675)	(0.0127)	3	-0.3146**	-0.1244
Immigrant				(0.1335)	(0.11628)
1	-0.14952**	-0.1199**	4	-0.04753	0.039204
	(0.0287)	(0.0250)		(0.1664)	(0.1445)
2	-0.1207**	-0.0824**	5	0.02841	0.04092**
	(0.0336)	(0.0293)		(0.3133)	(0.144)
Subsidized-lunch			Parents' educ	0.03336**	0.023155**
1	-0.1164**	-0.0836**		(0.0011)	(0.0009)
	(0.0351)	(0.022)	Average parents educ	-0.00133	0.01726
2	-0.1732**	-0.1545**		(0.0313)	(0.0026)
	(0.0332)	(0.0163)	S_j	-0.7500**	-0.0251**
Employment Father				(0.0562)	(0.009)
2	-0.0713**	-0.0733*	Year dummies	YES	YES
	(0.2055)	(0.0253)	R^2	0.2865	0.2468
3	0.12858	0.27321	Adj-R^2	0.2855	0.2456
	(0.2066)	(0.1799)	No. Obs.	13347	13347
4	0.01911	0.00252			
	(0.0773)	(0.0673)			
5	-0.0607**	-0.02			
	(0.0277)	(0.0241)			

Note: standard errors are reported in parentheses. ** - significant at 5%, * - significant at 10%