THE EVOLUTION OF LOCAL SCHOOLING SYSTEMS AT THE SECONDARY LEVEL IN BRAZIL BETWEEN 1991 AND 2010

Abstract

In spite of the negative overall representation of Brazilian secondary schools, the educational system at this level has shown remarkable quantitative developments. This paper proposes evolutionary paths of development of local secondary educational systems and addresses their determinants. It uses four schooling indicators - age specific enrollment ratio, gross enrollment ratio, net enrollment ratio, and age-grade distortion - conjointly with the use of cluster analysis to determine development trajectories. Furthermore, it analyses the main determinants of the evolutionary paths with the use of logistic and multinomial logistic models. Regional aspects related to socioeconomic level, labor market returns to human capital, demographics, social stability, location and microenvironment were associated with educational evolution. In addition, the evolution of the schooling system was associated with schooling performance.

Keywords: education, Brazil, secondary level, evolutionary paths, schooling performance

JEL I21, I25, R10

THE EVOLUTION OF LOCAL SCHOOLING SYSTEMS AT THE SECONDARY LEVEL IN BRAZIL BETWEEN 1991 AND 2010

André Braz Golgher

1 - Introduction

This most common perspective in referring to the Brazilian secondary educational system is quite negative: repetition and dropout rates are high and temporary leave is common (INEP, 2008). In addition, schooling performances are poor, even when compared to countries with similar per capita income (PISA, 2009). Consequently, the majority of Brazilian youth are unable to participate adequately in the labor market and/or to pursue higher levels of formal education (Neri et al, 2008), and the traditional model of secondary school in Brazil does not fulfill the expectations of most Brazilians (Castro et al, 2000; Corti and Souza, 2009). The youths' potential for income generation is poor and social mobility is limited (Hanushek and Wossmann, 2007; Schwartzman and Cossio, 2007). Associated with this is the recent increase of the 'no, no, no' generation among Brazilian youth: those who do not study, do not work, and do not look for a job (Camarano and Kanso, 2012).

This brief and bleak view emphasizes some of the problems faced by the Brazilian secondary schooling system. However, not neglecting its notable limitations, some remarkable recent advances were observed, especially in enrollment ratios and age-grade distortion.

This paper discusses the expansion of secondary schooling systems in municipalities in Brazil between 1991 and 2010 with the use of four schooling indicators: age specific enrollment ratio, gross enrollment ratio, net enrollment ratio, and age-grade distortion (Riani and Golgher, 2004; UNESCO, 2009). Similarly to many other studies (for instance, Barro and Lee, 2001; Goldin and Katz, 1997; Meyer et al, 2012; Pontili and Kassouf, 2008; Urquiola and Calderon, 2006), this paper discusses the expansion of the educational system using enrollment ratios and age-grade distortion; however, it takes a rather different approach.

It proposes evolutionary phases and paths of development of local secondary educational systems using these four indicators conjointly with the use of cluster analysis. The hypothesis is that regions evolve from a more primitive to a more advanced system following some preferential paths: some regions progress rapidly, while others lag behind; some areas primarily develop in one specific indicator, while others initially evolve in another. In addition, I link the findings related to the expansion of the secondary level schooling systems with academic performance. Some areas could greatly increase enrollment, decrease age-grade distortion, and still have a good average academic performance. That is, some regions could overcome potential trade-offs between quantity and quality. To the best of my knowledge, the use of four indicators conjointly describing the reality of Brazilian education was never pursued before.

In order to study the quantitative evolution of the secondary level educational system, I used as databases the Brazilian Censuses of 1991, 2000 and 2010, the three most recent at the time of this research. The data for schooling performance at the secondary level came from the National Exam of the Secondary Level (ENEM) of 2012, a short interval after the quantitative evolution observed by the Demographic Census.

This paper is structured in nine sections, including this introduction. The next section presents a literature review with topics associated with the empirical analysis of the paper. The third section presents the databases and the empirical strategy of the paper. The following section describes descriptive statistics for each of the four indicators cited above, emphasizing their evolution between 1991 and 2010. The fifth section characterizes regional profiles of educational systems, proposing developmental phases with the use of cluster analysis. The next section describes the preferential transitions between these phases, proposing theoretical and empirical paths for the modernization of regional educational systems. Section Seven describes the main determinants of these paths with the use of logistic and multinomial logistic models. The main objective is to understand why some regions show a more remarkable evolution through a particular path, while others present a more stagnant reality. Section Eight associates the findings of the previous sections with schooling performance. The objective is to link the expansion of educational systems with performance, emphasizing the regions that could evolve quantitatively and also show good levels of performance. The last section concludes the paper.

2 - The expansion of educational systems and schooling performance

Most individuals nowadays complete the elementary level in Brazil, the first nine years of formal education, which has become close to universalized (Lima, 2011). However, a sizable proportion of young Brazilians leave the educational system in the transition between the elementary and the secondary levels. The secondary level corresponds to the next three years of formal education.

Furthermore, dropout rates while attending the secondary level are quite high (Corti and Souza, 2009; INEP, 2008). Individuals argued that their main reasons to drop out of high school were lack of interest in the subjects taught and/or inability to cope with the extra demands of the secondary level (Corti and Souza, 2009; Neri et al, 2008; Schwartzman and Cossio, 2007).

Besides, there are high repetition rates at the secondary level in Brazil. Young Brazilians stated as the main reasons for their failure the following factors: lack of will, difficulty with the disciplines, difficulties balancing work and studies, and the obligation to look after brothers and sisters (Neri et al, 2008).

Even though these problems persist in the Brazilian educational system at the secondary level, there has been a notable increase in the numbers of high school graduates. For instance, the graduation rate increased from 22% to 50% for 20-year-old individuals between 1997 and 2007 (Golgher, 2010), though apparently the rate did not increase much afterwards.

Dropout and repetition rates directly affect many indicators that are commonly used to address regional heterogeneity of educational attainment. Among them are age specific enrollment ratio, gross enrollment ratio, net enrollment ratio, and age-grade distortion. In the following paragraphs, I present some studies that also used these or similar indicators. For detailed presentations of these indicators, see UNESCO (2009).

Meyer et al (2012) addressed the global expansion of mass education between 1870 and 1980. They stated that mass schooling became a worldwide institution, national educational expenditures increased, and nowadays even the poorest countries maintain a primary school system. The authors analyzed primary enrollment ratios of different countries. They observed that geographical characteristics, representing cultural, historical and institutional disparities, greatly influenced educational expansion, while urbanization, racial and religious composition affected the process less decisively. Therefore, they conclude that the rate of expansion was strongly shaped by structural location in global society.

According to Goldin and Katz (1997), the U.S. led the world in the expansion of secondary schooling; however, the evolution was not spatially homogenous. Some regions, such as those that had greater wealth, higher proportions of elderly people, greater income equality, lower participation of manufacturing in the job market, and lower proportions of migrants, had higher rates of secondary school enrollment and graduation.

Urquiola and Calderon (2006) compared countries in Latin America and the Caribbean using age specific enrollment ratios and schooling levels. The authors observed that most countries showed high enrollment rates, implying that students attended school for many years; however, these students displayed a great variation in educational attainments. In particular, Brazil had a relatively reasonably good indicator for enrollment, but was among the worst in effectiveness. The Brazilian youth attended the educational system for quite a long time vis-à-vis other countries in Latin America and Caribbean, but they did not acquire more education due to high school postponement and high rates of repetition.

Pontili and Kassouf (2008) analyzed the determinants of age-grade distortion in Brazil. They observed that individual factors, family characteristics, school infrastructure and location were associated with this indicator. They concluded that students from households with lower socioeconomic status would greatly benefit from school infrastructure improvements and from an increase in their social status.

Barro and Lee (2001) compared international data on educational attainment and observed that many previous studies used gross enrollment ratio and net enrollment ratio. The authors emphasized that such indicators do not directly measure human skills acquired in school, as quality of education might vary.

Thus, the negative perspective on the secondary level educational system in Brazil is not only represented by the high dropout and repetition rates, low enrollment ratios and high age-grade distortion, but also by the low levels of schooling performance (Soares, 2005; Machado et al, 2008). Many factors explain these low performance levels, such as personal features, household and family characteristics, location, school characteristics, and peer and contextual effects (Araújo and Siqueira, 2010; Barros et al, 2001; Duflo et al, 2011; Fernandes and Natenzon, 2003; Machado et al, 2008; Rodrigues et al, 2011; Rodrigues et al, 2013; Soares, 2005; Soares and Alves, 2013).

3 - Methodological section

This section presents in six subsections the methodology I applied in the paper. First, I present the databases. Then I detail the four selected indicators associated with enrollment and age-grade distortion. After this, I discuss how I defined the geographical units of analysis. Subsection 4 describes how the evolutionary phases were defined using cluster analysis. The empirical strategy of the paper is described in subsection 5. Subsection 6 describes the explanatory variables used to analyze the determinants of the evolutionary paths.

3.1 - Databases

The Demographic Censuses are done every ten years, are nationally representative, and can be applied to municipalities, the basic geographical unit of analysis of the paper. Moreover, they are suitable to address educational indicators such as those used here. The use of the last three Brazilian Demographic Censuses, those of 1991, 2000 and 2010, is justified by the necessity of observing the data in a time span of approximately two decades in order to evaluate evolutionary paths.

The data for school performance comes from the ENEM (national secondary level exam) of 2012, a short interval after the period determined by the Demographic Censuses. This exam was created in 1998 with the main objective of evaluating the performance of students at the end of the secondary level. Recently, since 2009, it is the main mechanism of selection for public tertiary education in Brazil. The exam has five sections: languages, mathematics, human sciences, natural sciences and writing skills. Results for subjects were highly correlated and a general mean was estimated across all exams.

3.2 – Educational indicators associated with enrollment and age-grade distortion

The use of stock variables, such as the average number of years of formal education of a population, are extremely useful to address points related to regional development; however, they represent the effects of cumulated educational attainment. Thus, the use of flow indicators tends to be more appropriate to analyze access to education, since these are more closely associated with the current characteristics of the educational system.

Among the flow variables, I selected four that can be estimated using census data and that complement each other: age specific enrollment ratio, gross enrollment ratio, net enrollment ratio, and age-grade distortion. I briefly present these indicators below. For a lengthier discussion see Unesco (2009) or Riani and Golgher (2004).

The age specific enrollment ratio is the percentage of individuals in a specific age group that attended school at any schooling level. The focus of the paper is to analyze the secondary schooling level in Brazil, and the ideal age range for attending this level is between 15 and 17 years old.

This indicator is particularly useful to give an overview of the retention capacity of the schooling system and to show the extent of educational participation of a specific age group (UNESCO, 2009). However, it has some ambiguities, especially for this age group (Riani and Golgher, 2004). Therefore, other indicators should complement this indicator while discussing regional heterogeneity of educational systems.

One such indicator is the gross enrollment ratio, which is the ratio between the number of individuals of any age who attend a specific schooling level and the number of individuals in the population in the age group eligible for the level. In this paper, I consider the ratio between all students at the secondary level, irrespective of age, divided by the population aged between 15 and 17. This indicator describes the capacity of the educational system to enroll students of a specific level, and determines whether a region has an educational system at this level which is adequate to incorporate all potential local students. This indicator has fewer ambiguities than the previous one when applied to the secondary level, but still presents some limitations (Riani and Golgher, 2004; UNESCO, 2009).

A third indicator, which complements these first two, is net enrollment ratio. This indicator is defined as the percentage of individuals of the official age for enrollment in a specific schooling level who actually attend the level. In this paper, that is the percentage of individuals aged between 15 and 17 that are studying at the secondary level. This indicator is appropriate to evaluate the efficiency of the schooling system, particularly concerning the flow of students, and it has fewer ambiguities than the previous two.

The last indicator addressed in this paper is age-grade distortion, which represents the proportion of students who are older than they should be in any schooling year--in this paper, in any of the three years of the secondary level. Similarly to the last indicator, the age-grade distortion is also associated with the efficiency of students´ progression. Like the first two indicators, this fourth indicator also has some ambiguities. For instance, if students who show age-grade distortion leave school, the indicator will improve, although real social conditions could be better if they had continued at school.

3.3 - Geographical units of analysis

This paper has the Brazilian municipalities as the basic geographical units of analysis. However, more than 1000 municipalities were created in the period: there were 4491 in 1991, 5507 in 2000, and 5565 in 2010. In order to make comparisons more insightful, I aggregated many municipalities in the last two Censuses and some in the first, and I obtained the same areas in all years. These areas will henceforth be called minimum areas of comparison (MAC). Most of them represented one municipality in 1991, but some are groups of two or more municipalities in this year with a total of 4253 areas.

3.4 The phases of development

This subsection describes how I defined the evolutionary phases of educational systems. Initially, I created the MACs for the period 1991 to 2010 and obtained the value for each indicator in each area. The data for the four indicators for each region were then pooled. Thus, initially the data consisted of 4253 MACs x 3 years for each indicator. Because the indicators have distributions with different means and standard deviations, and this might affect the clustering process, I obtained a rank for each indicator. With the four ranks, from one to 12759, representing from the best to the worst value, I classified all MACs/Years in homogenous groups with the use of cluster analyses.

Clustering methods are exploratory statistical procedures that classify the elements of information sets in internally homogeneous groups in order to find structure in data. There are different procedures and algorithms for clustering data, each with specific limitations and strengths (Hair et al., 2009; Jain, 2010). Given the nature of our variables and the size of our dataset, I chose to use the K-means clustering procedure. However, this technique has the drawback of requiring a previously defined number of clusters. I performed different analyses with a range of arbitrary defined numbers of clusters and, due to empirical outcomes, I chose the study with eight clusters. Finally, I classified the eight clusters in five phases of development, as detailed in the discussion of the results.

After obtaining the phases of development, I described the main transitions between them. Many areas initially classified as the least developed regions evolved to become some of the most developed areas, while others lagged behind.

3.5 - Econometric models

I used econometric models to verify which socioeconomic, demographic and geographic characteristics were significantly associated with this diversity of evolutionary paths. The empirical analysis is divided in two parts: the first discusses the main evolutionary paths between 1991 and 2000, and the second addresses those between 2000 and 2010. The dependent variable is binary or categorical and is the final cluster membership in each period conditioned in the profile of origin. For instance, regions classified as cluster 1 in 1991 mostly developed to cluster 3 or 4 in 2000. Then I created a binary variable for the cluster achieved in 2000 (1 – For cluster 4, 0 – For cluster 3), and used logistic models. The objective is to analyze factors associated with the different likelihoods of evolving from cluster one to four when compared to progressing from cluster one to three. Similarly, regions categorized as cluster 2 in 1991 mostly evolved to clusters 3, 5, 6 or 7 in 2000. Thus, I selected the regions classified in cluster 2 and created a categorical variable with the cluster achieved in 2000. The reference is always the cluster with the smaller number, representing the least remarkable evolution. For categorical dependent variables, I used multinomial logistics models. The same procedure was used for the second group of models that analyzed the steps between 2000 and 2010, and were conditioned on the cluster achieved in 2000. I will present more details about the dependent variable when describing the empirical results of the models.

3.6 – Explanatory variables

The independent variables of the econometric models consist of socioeconomic, demographic, geographical and political characteristics that might have an impact on educational indicators. Similar sets of explanatory variables were used by Goldin and Katz (1997) and by Meyer et al (2012). Most explanatory variables were obtained for the beginning of the period being analyzed to minimize endogeneity problems. Therefore, the variables are from 1991 and 2000 when not specified.

For urban structure, I included the logarithm of population and an urbanization index was built using principal components analysis (PCA) (Hair et al., 2009). A third independent variable consists of a socioeconomic index, also built using PCA. These two indexes were normalized means of the variables composing the index.

Regarding local economic activity and labor market returns on human capital, the models incorporated data for university enrollment, representing the return on a high school diploma. Moreover, the models included the proportion of migrants in the population. A migrant is someone who moved to a municipality in the last five years. High proportions of migrants are positively associated with positive characteristics of local labor markets; however, they can also be related to a lack of social stability. Other variables related to social stability included in the models are local unemployment rates and homicide rates. The data for homicide rates is a mean value of five years around the years 1995 and 2005, indicating the values in the middle of each decade.

Demographics may influence educational outcomes, in particular the population's age structure and sex distribution. The models include the proportion of individuals aged under 15 in the population. Higher values for the proportion of this age group may indicate greater pressures to incorporate larger numbers of students in the educational system. Sex distribution, measured by sex ratios, was also included in the models. Girls in Brazil tend to achieve higher levels of formal education (Haussmann and Golgher, 2016).

Geography might also influence the development of the educational system. A dummy was created indicating whether or not the area was located in the North or Northeast regions, the two least developed in Brazil. The reference was if the area was localized in one of the most developed macroregions in Brazil: the Central-West, South or Southeast regions.

Homicide rates might represent social instability or may be spuriously correlated with positive features of a society depending on the level of development of a region. I included an interaction term between these rates and the dummy representing regions.

Moreover, spatial heterogeneity and/or dependence among the geographical units of analysis might exist (Elhorst, 2010; LeSage and Pace, 2009). For instance, there are some evidences of interactions between regions and educational inequalities (Tselios, 2008). In order to represent the potential influence of the microenvironment of the region, the neighbors' mean values of the four above-mentioned educational indicators were initially included in the models. After some analyses, the spatial lagged values of age specific enrollment ratio and of net enrollment ratio were dropped from the models.

Although the models were conditioned on the cluster of origin, and they represent approximately homogeneous groups, regions in each cluster were not identical. Hence, as a control, the models included the initial values for the four educational indicators.

I included two other variables for the analysis of the period between 2000 and 2010. Notice that in this second group of analysis, the areas are conditioned on the cluster achieved in 2000, and the regions classified as cluster 1 or 2 in 1991 are analyzed together. Thus, the models include a dummy indicating whether the region was classified in cluster one in 1991.

The Brazilian Federal Government recently implemented the Bolsa Família program, a cash transfer policy that might affect school attendance. This program assists close to 25% of the Brazilian population and was initially designed to promote education (Fiszbein and Schady, 2009). The proportion of households assisted by this program in each area in 2004 was included in the models.

4- Evolution of schooling indicators in Brazil between 1991 and 2010

This section presents an overview of the evolution of the four selected educational indicators for MACs in Brazil in the period between 1991 and 2010. Table 1 shows the proportion of areas that improved in each indicator in each period. Nearly all areas improved in the four indicators between 1991 and 2010, with values ranging from 97% of the total for age-grade distortion to 100% for net enrollment ratio. This indicates that despite the negative overall representation of Brazilian secondary schools, the educational system at this level exhibited remarkable quantitative developments over this period.

From 1991 to 2000, similar numbers were observed for the enrollment indicators, but not for age-grade distortion. The results for age specific enrollment ratio, gross enrollment ratio and net enrollment ratio respectively indicate that the proportion of the population aged 15 to 17 in school increased, more students were at the secondary level, and more students aged 15 to 17 were at the secondary level. However, given the results of the age-grade distortion indicator, many older students that ideally should be at the tertiary level were still in high school, or many were in the secondary level, but at a lower level than they should ideally have been.

Between 2000 and 2010, the proportion of individuals in the population aged 15 to 17 in school did not increase so remarkably, nor did the number of older students at the secondary level, as shown by the not so notable evolution of age specific enrollment ratio and gross enrollment ratio indicators. However, net enrollment ratio and age-grade distortion improved the most, indicating the increase in efficiency of students´ progression.

This brief analysis suggests that the indicators do not evolve at the same pace: the age specific enrollment ratio and gross enrollment ratio apparently increase before the other two. That is, first retention increases, then there is a correction in the students' progression.

Table 1 – Proportion of areas of comparison with positive evolution in Brazil

Indicator	**Period**		
	1991 and 2000	2000 and 2010	1991 and 2010
Age specific enrollment ratio	99.2	87.9	99.8
Gross enrollment ratio	99.4	86.3	99.8
Net enrollment ratio	97.9	94.9	100.0
Age-grade distortion	70.8	98.4	97.0

Source: FIBGE, 1991, 2000 and 2010

5 - Evolutionary phases of local educational systems

In the previous section, I discussed the evolution of the four indicators separately, when I observed that the temporal dynamics of the indicators differed. This section analyzes the joint dynamics of all four indicators. The questions being addressed here is: How do regions evolve from a less developed local schooling system to a more developed one?

Table 2 shows the results of the clustering process in two panels. The upper panel shows the mean value for the ranks of each indicator for each cluster centroid. Notice that the values could potentially vary from one to 12759: the smaller the value, the better the indicator. The results should be analyzed relative to the other clusters. Additionally, the table includes a brief description of each profile in the lower panel in order to facilitate the comprehension of the main characteristics of the clusters.

As already mentioned, I classified the eight clusters into five phases of development and they were numbered according to these phases; the first cluster characterized the least developed regions, while the last represented the most developed areas. In the next paragraphs, I clarify how I defined these phases.

Cluster 1 characterized the regions with very low levels for all three indicators of enrollment and very high age-grade distortion when compared to other clusters. Notice that relative to the other clusters, the values for the cluster centroid are among the highest for all four indicators, all above 10,000. Thus, the cluster categorized areas with very few individuals aged 15 to 17 studying, very few secondary level students, and very low efficiency of students' progression. Given these characteristics, it was the only cluster defined as phase one and was named the "Least developed areas." These features described 2,175 MACs/Years: in 1991, 1,988 MACs were in cluster 1, while in 2000 only 187 MACs belonged to this cluster. Therefore, this profile was a typical profile of 1991, as very few regions had these characteristics in 2000 and none in 2010.

I considered the clusters numbered 2, 3 and 4 as phase two. They are clearly and unambiguously more developed profiles than cluster one. Notice, however, that these three clusters classified as phase two show different development profiles. Each specific cluster has better values for some indicators and worse values for others when compared with the other clusters in the same phase of development, and no cluster is unequivocally more developed than the others.

Cluster 2 is similar to cluster 1 in the two first enrollment indicators, as the values for the centroid are similar. However, as shown by the results of net enrollment ratio and age-grade distortion, with smaller centroid values for both indicators, students´ progression was more efficient. Hence, it can be considered as a potential initial step of an evolutionary path from the least developed regions with a correction in the students' progression. The profile was named "least developed areas with lower age-grade distortion." Cluster 2 categorized more than 1000 MACs/Year, mostly also from 1991.

Cluster 3 is a typical profile of 2000 and characterized over 1000 MACs/Years. It mostly differs from cluster 1 in two indicators: the age specific enrollment ratio and the gross enrollment ratio, both higher. These differences suggest an evolution from the characteristics of cluster 1 with an increase in the number of students aged 15 to 17 at any schooling level and an increase of older students at the secondary level. However, students' progression continues to be inefficient, with low values for net enrollment ratio and very high values for age-grade distortion. I named this cluster "less developed areas with higher age-grade distortion."

Cluster 4, named "less developed areas with lower age-grade distortion," characterized 1,264 MACs/Years, mostly from 1991 and 2000. It is a more advanced development profile than that observed in cluster 1 with a slight evolution in all indicators: a greater proportion of students among the population aged 15 to 17, more students at the secondary level, more students at the secondary level aged 15 to 17, and less age-grade distortion. Notice that this cluster represents a general development when compared to cluster 1, unlike clusters 2 and 3, which show improvements in a limited set of indicators. However, cluster 4 was not considered a development from clusters 2 or 3, as some indicators are less developed in cluster 4 than in these two other clusters.

I classified cluster 5 as phase three, because it is a clear development from clusters 2 and 4 in phase two. Comparing clusters 2 and 5, there was an increase in all enrollment indicators and a decrease in age-grade distortion. A comparison between clusters 4 and 5 shows the same low values for age specific enrollment ratio; however, there was an increase in gross and net enrollment ratios and a decrease in age-grade distortion.

Finally, comparisons between clusters 3 and 5 do not show such a clear dominance of the latter over the former as the last two mentioned comparisons. Regions classified in cluster 5 had larger proportions of individuals in the secondary level, with less age-grade distortion, when compared to the areas categorized by cluster three; however, the proportion of individuals aged 15 to 17 in school was smaller. Nonetheless, given the ambiguities of the age specific enrollment ratio indicator, cluster 5 might be considered a potential development from cluster 3: due to an increase in the students' progression efficiency, dropout rates among individuals of the age group between 15 and 17 might increase as the student finishes the elementary level. I characterized Cluster 5 as "intermediate level of development areas," and it classified regions in all three years.

Clusters 6 and 7 represented the fourth phase of development, and they show a more developed profile than all clusters in the previous phases. Notice that both profiles mostly characterized MACs in 2000 and in 2010. The main difference between clusters 6 and 7 is the proportion of students aged 15 to 17 that are still in the elementary level, which is greater in the first, with higher levels of age-grade distortion. Clusters 6 and 7 were respectively named “highly developed areas with high age-grade distortion” and “highly developed areas with low age-grade distortion.”

The evolution from cluster 5 to cluster 6 indicates an increase in the proportion of students aged 15 to 17, and of older students at the secondary level, with slight increases in age-grade distortion and in net enrollment ratios. That is, although age-grade distortion slightly increased, this was caused mostly because older students did not drop out of the educational system. From cluster 5 to cluster 7, the evolution is a general development in all indicators.

The last cluster had good levels for all indicators, clearly dominated all others, and was defined as phase five: the “most developed areas.” This cluster grouped 2269 MACs/Years, most in 2010. The characteristics are a natural adjustment of students’ progressions from clusters six and seven with an increase in the proportion of individuals aged 15 to 17 attending school.

Table 2 – Phases of development of local educational systems obtained by cluster analyses

Cluster	Phase	Type	Indicator				MACs			
			ASER	GER	NER	AGD	Total	Proportion in 1991	Proportion in 2000	Proportion in 2010
1	1	Unique	10648	11322	11421	10951	2175	91	9	0
2	2	A	11183	10153	9162	5862	1272	96	4	0
3	2	B	5249	8379	9525	10724	1162	7	91	2
4	2	C	8081	7623	7852	8811	1264	44	53	3
5	3	Unique	7990	6409	5723	4877	1249	31	53	17
6	4	A	2911	3525	4994	6829	1681	1	40	59
7	4	B	5166	3912	3121	2549	1687	0	33	67
8	5	Unique	1814	1746	1369	2054	2269	0	18	82
Main regional features										
1	Least developed areas									
2	Least developed areas with lower age-grade distortion									
3	Less developed areas with higher age-grade distortion									
4	Less developed areas with lower age-grade distortion									
5	Intermediate level of development areas									
6	More developed areas with high age-grade distortion									
7	More developed areas with low age-grade distortion									
8	Most developed areas									

Source: FIBGE, 1991, 2000 and 2010
ASER = Age specific enrollment ratio
GER = Gross enrollment ratio
NER = net enrollment ratio
AGD = Age-grade distortion

6 - Evolutionary empirical paths of local educational systems

The discussion above presented some potential and theoretical steps in the evolution of educational systems. However, many steps might not be common, as some clusters typically characterized profiles in specific years. This section presents the most frequent evolutionary steps.

I begin the discussion with four clusters that represented most areas in 1991, those numbered 1, 2, 4 and 5. Very few areas in 1991 had the characteristics associated with the other clusters. Moreover, the regions classified in clusters 4 and 5 in 1991 were much less numerous than those classified in clusters 1 or 2, and their evolution was not so remarkable or diversified. The areas were already more developed in the beginning of the analysis, and most evolved to cluster 8 in 2010. Therefore, the following analysis focuses on the regions that were classified in clusters 1 or 2 in 1991, which represent 3,211 areas out of 4,253.

Table 3 shows the empirical evolution of the areas categorized in clusters 1 or 2 in 1991 between 1991 and 2000. Cluster 1 categorized 1,988 regions in 1991, close to half of all areas in Brazil. Approximately 70% of these regions followed only two paths: 23.3% made the transition to cluster 3, and 45.6% evolved to cluster 4. These more common evolutionary paths respectively evolved from the "least developed areas" to "less developed areas with higher age-grade distortion" or to "less developed areas with lower age-grade distortion." In the first path, there was a more remarkable increase in the proportion of individuals aged 15 to 17 who were studying; however, many of these were still attending the elementary level. In the second path there was an increase in the proportion of individuals aged 15 to 17 who were studying as well as an increase in the number of secondary level students and a decrease in age-grade distortion. That is, all indicators improved conjointly, an apparently more robust development than the previous path. Concerning the other transitions, some regions, 8.7% of the total, did not develop at all, while around 20% evolved more remarkably than in these more common paths.

Cluster 2 classified 1,223 areas in 1991. More than 80% of them followed four paths between this year and 2000: 13.2% evolved to cluster 3, 32.2% developed to cluster 5, 13.4% to cluster 6, and 24.4% to cluster 7. The first of these paths, from "least developed areas with lower age-grade distortion" to "less developed areas with higher age-grade distortion," showed an increase in the number of students aged 15 to 17 in all levels, with an increase in age-grade distortion. In the path between the "least developed areas with lower age-grade distortion" to "intermediate level of development areas,” there was a more remarkable increase in the proportion of students in general, not accompanied by increasing age-grade distortion. Furthermore, some regions evolved from “least developed areas with lower age-grade distortion” to “highly developed areas with high age-grade distortion,” while others evolved from this first cluster to “highly developed areas with low age-grade distortion.” In both cases, the development was substantial, and in this first step the increase in the proportion of students aged 15 to 17 in the elementary level was greater than in the latter step. Regarding the other paths, close to 14% evolved to clusters 4 or 8, and 2.4% did not apparently develop. Finally, notice in the last columns that regions classified in 1991 as cluster 3 to 8 mostly evolved to cluster 6 to 8.

Table 3 – Transitions between 1991 and 2000 for areas classified as cluster 1 or 2 in 1991

		Cluster in 1991			
		Least developed areas	Least developed areas with lower age-grade distortion	1 and 2	3 to 8
Cluster in 2000 (%)	Least developed areas	8.7	1.0	5.8	0.0
	Least developed areas with lower age-grade distortion	1.6	1.4	1.5	0.0
	Less developed areas with higher age-grade distortion	23.3	13.2	19.4	5.0
	Less developed areas with lower age-grade distortion	45.6	7.2	31.0	5.7
	Intermediate level of development areas	9.6	32.2	18.2	6.9
	More developed areas with high age-grade distortion	9.1	13.4	10.7	31.9
	More developed areas with low age-grade distortion	1.5	24.4	10.2	21.8
	Most developed areas	0.7	7.2	3.1	28.8
	Total	1988	1223	3211	1042

Source: FIBGE, 1991 and 2000

Similarly to table 3, table 4 shows the evolution between 2000 and 2010. First, notice that the large majority of regions, 2,877 out of 3,211, were classified in clusters 3 to 7 in 2000. Most of them achieved one of the three most developed profiles. The MACs classified in 2000 as “less developed areas with higher age-grade distortion” or “less developed areas with lower age-grade distortion” mostly evolved to become highly developed areas or the most developed areas, with a general increase in the number of students at the secondary level and a correction of student progression. Regions that were already more developed in 2000, in clusters 5, 6 or 7, mostly achieved one of the two last clusters, also with a general improvement of all indicators.

Table 4 – Transitions between 2000 and 20102000 for areas classified as cluster 1 or 2 in 1991

		Cluster in 2000							
		1 or 2	3	4	5	6	7	8	Total
Cluster in 2010 (%)	1 to 5	38.2	8.7	8.5	3.9	2.3	0.0	1.0	8.1
	More developed areas with high age-grade distortion	43.8	24.7	61.8	1.4	8.4	0.6	0.0	28.3
	More developed areas with low age-grade distortion	15.0	42.0	16.8	52.6	21.5	34.7	14.9	30.4
	Most developed areas	3.0	24.7	12.9	42.1	67.7	64.7	84.2	33.2
	Total	233	624	995	585	344	329	101	3211

Source: FIBGE, 2000 and 2010

The results above suggest that empirical evolutionary paths depend on the velocity of changes. On the one hand, for quicker evolutions, such as those observed in the 90s in Brazil, most regions begin their development of the secondary level with an increase in the retention capacity of students aged 15 to 17 and an increase in age-grade distortion. Then the gross enrollment ratio increases, as the retention capacity of older students is also improved. After this, students' progression becomes more efficient, with an increase in net enrollment ratio and a decrease in age-grade distortion. A more homogeneous path is observed if there is a slower evolution in the indicators, as verified in the 80s in Brazil. Because of the slower joint evolution of the indicators, I did not observe a very large enrollment ratio for those aged 15 to 17 with high levels of age-grade distortion before the other indicators progress.

7 - Empirical analyses of the evolutionary paths

The last section emphasized that many areas initially classified in clusters 1 and 2 in 1991 fully developed to cluster 8 in 2010, while others regions still lagged behind, mostly in clusters 6 and 7. This section addresses the question of which features might be associated with this diversity in evolutionary paths.

Table 5 presents the results of the econometric models for the most numerous transitions between 1991 and 2000. The first model is a logistic model and describes the results for the regions that were classified in 1991 as the "least developed areas." As already discussed, most regions made the transition in 2000 to the clusters "less developed areas with higher age-grade distortion" or "less developed areas with lower age-grade distortion." The model compares these two outcomes, and the reference is the first cluster. That is, it analyses factors correlated with the initial development steps of local educational systems with or without a greater efficiency of the students' progression.

Most values were significant at 5%. Notice that controls for the initial values of the four indicators were included in the model.

Concerning urban structure, more populated municipalities, possibly due to larger schools, had a smaller propensity of belonging to the cluster with lower age-grade distortion. The variable for urban index was not significant in this model nor in the next. Regions that had a better socioeconomic level, and a larger proportion of university enrollments, indicating greater returns on human capital, also had a greater propensity of belonging to the cluster with lower age-grade distortion.

Regarding social stability, areas with lower unemployment rates had a greater propensity of belonging to the cluster with lower age-grade distortion. The variable for homicide rate was not significant in this model nor in the next. As described earlier, a higher proportion of migrants might indicate a local booming economy, but can also imply a higher level of social instability. The results indicate a positive correlation with the evolution of educational systems, suggesting that the first factor may be more decisive.

Areas with greater proportions of males tended to develop more, suggesting that low levels of educational development favor men, possibly in part due to less sophisticated labor markets, lower levels of formal education and greater gender polarization. The variable for the proportion of the population aged 0 to 14, proxy for population pressure on the educational system, was not significant in this model nor in the next.

Finally, location matters. The extremely small odds-ratio for the dummy North/Northeast regions indicate that macrolocation is greatly associated with the evolution of students' progression, possibly due to cultural factors, associations with local labor markets and differences in states' policies and priorities. Regions with low levels of educational development that are located in more advanced states and macroregions tend to develop more remarkably. Besides, the microenvironment is also related to corrections in students' progression, as shown by the value below one for the odds-ratio for the lagged age-grade distortion. Municipalities surrounded by regions with lower age-grade distortion also tended to show lower age-grade distortion, suggesting localized coevolution. The interaction between homicide rates and macroregions was non-significant in both models. The lagged value for gross enrollment ratio was not significant in the first model.

The next model has the same explanatory variables but compares the areas that were classified as “least developed areas with lower age-grade distortion” in 1991 in a multinomial model. The reference is again the cluster “less developed areas with higher age-grade distortion,” which is compared to three other profiles: areas with an intermediate level of development, or more developed areas with high or low age-grade distortion.

Similarly to what was observed in the previous model, regions with better socioeconomic levels, greater proportions of enrollments in the tertiary level and of migrants, lower unemployment rates, and locations outside the North or Northeast regions tended to develop their educational systems more fully. But in this model, unlike the previous one, greater proportions of females were positively related to a greater development of the educational system. These results suggest that for regions with a very incipient schooling system, as in the first comparison, greater proportions of males tend to promote a greater development until a reasonable level is attained; after this females take over, possibly due to lower participation in the labor market and because of cultural evolution towards women’s empowerment (Hausmann and Golgher, 2016). Concerning the microenvironment, municipalities encircled by areas with greater proportions of students in the secondary level and with lower levels of age-grade distortion had a greater propensity to develop more. That is, local educational systems are influenced by what happens in their vicinities. The size of local population was significant at low to intermediate levels of development, but not in the other comparisons.

Table 5 – Logistic and multinomial logistic models for preferential paths between 1991 and 2000

Model:	Logistic	Multinomial logistic		
Cluster in 1991:	Least developed areas	Least developed areas with lower age-grade distortion		
Comparison with:	Less developed areas with lower age-grade distortion	Intermediate level of development areas	More developed areas with high age-grade distortion	More developed areas with low age-grade distortion
Reference:	Low developed areas with higher age-grade distortion			
Variables	Odds-ratio			
Logarithm of population	0.84*	1.38*	0.86	1.37
Urban index	0.99	1.00	1.01	0.99
Socioeconomic index	1.05*	1.04*	1.06*	1.10*
University enrollment	3.08*	1.94	2.06	3.01*
Unemployment rate	0.94*	0.89*	1.01	0.77*
Homicide rate	0.98	0.99	1.01	0.97
Proportion of migrants	1.04*	1.04*	1.05*	1.05*
Sex ratio	1.04*	0.98	0.93*	0.90*
0 to 14 years old	0.98	1.02	1.01	1.07
North or Northeast regions	0.08*	0.24*	0.39	0.05*
Homicide*region	1.04	1.00	0.95	0.92
Lagged gross enrollment ratio	1.01	1.06*	1.06*	1.12*
Lagged age-grade distortion	0.97*	0.94*	0.98	0.93*

Controls for initial values of
age specific enrollment ratio, gross enrollment ratio, net enrollment ratio and age-grade distortion
Note: * significant at 5%
Source: FIBGE, 1991, 2000 and 2010

Table 6 shows the results for the transition between 2000 and 2010 for the five clusters that were particularly common in 2000. Note that the models are conditioned on the cluster achieved in 2000. However, only those regions classified in 1991 as cluster 1 or 2 were included in the models. Again, controls for the initial value of the four indicators are included in the models.

Before discussing any particular model, I present some general findings. Notice in the last line of the table that most coefficients for the dummy indicating the cluster of origin in 1991 were non-significant. That is, after controlling for initial values of age specific enrollment ratio, gross enrollment ratio, net enrollment ratio and age-grade distortion, the origin in 1991 did not account significantly for the posterior trajectories between 2000 and 2010 in most cases. The origin in 1991 was significant only in the first model and with odds-ratio below one, indicating the lower propensity of regions that originated in this profile to develop more fully.

Besides, all the coefficients for Bolsa Família program were non-significant, indicating no significant correlation between this policy and the evolution of educational systems at the municipal level. Although I controlled for many socioeconomic factors, it should be clear that the least developed regions tend to have a greater proportion of their households as recipients of this conditional cash transfer policy. Behrman, Parker, and Todd (2005) and Fiszbein and Schady (2009) verified evidence that this type of policy increases schooling attainment of individuals.

Now I focus the discussion on each of the five models. Among those classified in 2000 in the profile “less developed areas with higher age-grade distortion,” most regions attained three clusters in 2010: “more developed areas with high age-grade distortion,” “more developed areas with low age-grade distortion,” and “most developed areas.”

Few variables were significant at 5% and we present their results. Areas with greater proportions of students at the tertiary level and a more favorable microenvironment showed a greater propensity for major improvements in student progression or development of the education system in general. Urbanization and social stability were also slightly associated with these greater improvements. The other variables showed non-significant coefficients, including the dummy for macroregion.

The next model addresses those areas that were classified in 2000 as “less developed areas with lower age-grade distortion.” These areas had a lower age-grade distortion in 2000 than those addressed in the previous models; however, they had a similar evolution.

The coefficients for urban structure, for the socioeconomic index and for university enrollment were non-significant. Similarly to the comparisons in previous table, areas with a higher proportion of immigrants, a location outside the North/Northeast regions, and a favorable microenvironment showed a greater likelihood of remarkably improving student progression or overall development of their education systems.

Concerning social stability, the lower the homicide rate, the greater the evolution of the educational system for areas located in the South/Southeast/Central-West regions. In the North/Northeast, as shown by the interaction term, the tendency is blurred. These results suggest that homicide rates in more developed areas are more associated with social stability, influencing educational outcomes, while in less developed ones this is not so straightforward.

Regarding the demographic variables, a higher proportion of females was related to lower levels of age-grade distortion, but not to greater general development. Females tend to be more focused on education (Samantha and Golgher, 2016), resulting in lower dropout and repetition rates. Lower population pressure, represented by smaller proportions of individuals aged 0 to 14 in the population, was associated with a greater development of the educational system.

The last three models address those regions that were classified in 2000 respectively as “intermediate level of development areas,” “more developed areas with high age-grade distortion” or “more developed areas with low age-grade distortion,” the most developed main profiles in 2000. These areas mostly evolved to two profiles in 2010, which are this last one mentioned and the “most developed areas.” The models compare these two profiles with logistic models analyzing the factors associated with a more remarkable educational development of the already most developed areas in 2000.

All coefficients for urban index, socioeconomic index, tertiary level enrollment and proportion of migrants were non-significant. Differently from most other comparisons, macro location was not associated with the evolution of the educational systems. This suggests that more developed areas in less developed macroregions more closely resemble those in more developed regions.

Similarly to the previous comparisons, two measures of social stability were significantly associated with the observed differences: homicide rates and unemployment rates. Lower population pressure was related to greater evolution, as in the previously discussed model. The microenvironment also seems to matter, as observed in previous comparisons. Similarly, regions surrounded by areas with higher gross enrollment ratio evolved more. Differently from the previous comparisons, however, those regions bordered by areas with higher age-grade distortion also evolved more. This seems contradictory; nonetheless, the regions analyzed here were already more developed in 2000, and hence age-grade distortion might occur at higher levels and continue for fewer years. Finally, in the last model, areas with a higher proportion of males evolved more, suggesting that more advanced labor markets interact with schooling systems favoring men.

Table 6 – Logistic and multinomial logistic models for preferential paths between 2000 and 2010

Model:	**Multinomial logistic**		**Multinomial logistic**		**Logistic**	**Logistic**	**Logistic**
Cluster in 2000:	**Less developed areas with higher age-grade distortion**		**Less developed areas with lower age-grade distortion**		**Inter-mediate level indicators**	**More developed areas with high age-grade distortion**	**More developed areas with low age-grade distortion**
Comparison with:	**More developed areas with low age-grade distortion**	**Most developed areas**	**More developed areas with low age-grade distortion**	**Most developed areas**	**Most developed areas**		
Reference:	More developed areas with high age-grade distortion				More developed areas with low age-grade distortion		
Variables	Odds-ratio						
Log. of population	0.87	0.91	1.13	1.36	0.87	1.48*	1.25
Urban index	0.99	0.97*	0.98	1.00	0.99	0.99	1.01
Socioeconomic index	0.99	1.04	1.00	1.02	0.99	1.02	0.98
University enroll.	2.24*	1.72	1.87	1.52	1.06	1.00	1.37
Unemployment rate	0.97*	0.98	0.98	0.98	0.99	1.03	0.92*
Homicide rate	1.16	1.02	0.93*	0.89*	0.98	0.97*	0.99
Prop. of migrants	0.97	0.97	1.06*	1.06*	0.98	0.99	1.00
Sex ratio	1.00	1.02	0.93*	0.97	1.03	1.03	1.09*
0 to 14 years old	0.99	0.99	0.99	0.92*	0.92*	0.89*	0.92
North/Northeast reg.	1.97	2.73	0.41*	0.25*	0.66	0.60	0.00
Homicide*region	1.16	1.02	1.06*	1.09*	1.04	1.02	17.36
Lagged gross enrollment ratio	1.03*	1.02*	1.05*	1.05*	1.02*	1.01	1.03*
Lagged age-grade distortion	0.88*	0.84*	0.95*	0.99	1.03	1.05*	1.06*
Bolsa Família prog.	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cluster 1 in 1991	0.56*	0.60	0.74	0.69	1.23	0.68	1.03
Controls for initial values of age specific enrollment ratio, gross enrollment ratio, net enrollment ratio and age-grade distortion							

Note: * significant at 5%

Source: FIBGE, 1991, 2000 and 2010

.

8 - Overcoming tradeoffs in local schooling systems at the secondary level in Brazil

This section associates the expansion of educational systems with school performance in a very brief discussion. The objective is to observe the potential trade-offs between quantity and quality, and to point out some regions that could expand their educational systems and achieve reasonable levels of performance.

Table 7 focuses on the regions that fully developed from cluster 1 in 1991 to cluster 8 in 2010. Only regions above a given threshold of schooling performance are shown. Among these regions, 96 areas had a mean performance of over 500, 60 regions had a value over 510, and 26 had a value over 520. These regions were located in five of the 27 states in Brazil, in particular in the state of Minas Gerais,. That is, location matters considerably.

Table 7 – Number of regions that fully developed between 1991 and 2010 above schooling performance thresholds

State	Above 500	Above 510	Above 520
Minas Gerais	83	54	25
Rio de Janeiro	3	2	0
São Paulo	7	3	1
Paraná	1	0	0
Goiás	2	1	0
Total	96	60	26

Source: FIBGE, 1991, 2000 and 2010. ENEM, 2012

9 - Conclusion

This paper proposed local evolutionary paths of development for secondary schooling systems in Brazil with the use of four indicators: age specific enrollment ratio, gross enrollment ratio, net enrollment ratio, and age-grade distortion. These indicators possess some ambiguities and limitations, which might sometimes render the separate analysis of each of them misleading for social researchers and policy makers. I defined evolutionary paths by the conjoint analysis of these four indicators, thus overcoming many of the limitations of more specific discussions focused on one or some of these indicators separately, and facilitating future hypotheses for the development of local educational systems.

Initially, I defined eight profiles, from the least developed to the most developed areas, for the municipalities in Brazil in the period between 1991 and 2010, using cluster analysis. Then, I observed the main empirical evolutionary paths of local educational development. Finally, I addressed some of the determinants associated with the diversity of paths with the use of logistic and multinomial logistic models.

Here are some of the main factors associated with differences in the evolutionary paths: For localities with low levels of educational development, macrolocation matters, possibly due to cultural and regional factors which are not as relevant in more developed areas. Besides, microenvironment matters, possibly because of localized regional cultures or comparison between close localities. Local socioeconomic levels were more positively associated with development in the initial stages of educational development, and social stability was positively related to evolution for all levels of evolution. Demographic factors might also affect development paths differently depending on the level achieved by the region. Population pressure was more associated with more developed evolutionary steps, while greater proportions of females were more associated with intermediary steps of evolution. Urban structure did not show a significant relation with educational evolution, nor did national policies, such as Bolsa Família.

The results observed in the paper indicate how local schooling systems evolve. This insight may facilitate the implementation of policies at the municipal and state levels designed to promote more advanced and encompassing educational systems. The paper's findings are also linked to students' performance, showing some areas that could overcome some of the trade-offs between quantity and quality. Which are the main determinants that enabled some areas to develop their secondary educational systems while obtaining reasonable levels of schooling performance? This is a venue for future research.

References

Araújo, F. and Siqueira, L. (2010). Determinantes do desempenho escolar dos alunos da 4ª. série do ensino fundamental no Brasil. *Economia e Desenvolvimento*, 9 (1), 70-103.

Barro, R. and Lee, J. (2001). International data on educational attainment: updates and implications. *Oxford Economic Papers*, 3, 541-563.

Barros, R. P., Mendonça, R., Quintães, G. and Santos, D. (2001). Determinantes do desempenho educacional do Brasil. *Pesquisa e Planejamento Econômico*, 31 (1), 1-42.

Behrman, J., Parker, S. and Todd, P. (2005). *Long-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico* (Discussion Paper No. 122). Ibero-America Institute for Economic Research, Göttingen, Germany.

Camarano, A. and Kanso, S. (2012). O que estão fazendo os jovens que não estudam, não trabalham e não procuram trabalho? *Boletim Mercado de Trabalho: conjuntura e análise*, 53, 37-44.

Castro, C., Carnoy, M. and Wolff, L. (2000). *Secondary schools and the transition to work in Latin American and the Caribbean* (Sustainable Development Department Technical Papers Series). Inter-American Development Bank.

Corti, A. and Souza, R. (2009). *Que Ensino Médio queremos*? São Paulo: Ação Educativa.

Duflo, E., Dupas, P. and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101 (5), 1739-1774.

Elhorst, J. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, 5 (1), 9-28.

Fernandes, R. and Natenzon, P. (2003). A evolucao recente do rendimento escolar das criancas brasileiras: uma reavaliacao dos dados do Saeb. *Estudos em Avaliação Educacional*, 28, 3-22.

Fiszbein, A. and Schady, N. (2009). *Conditional cash transfers reducing present and future poverty*. Washington: A World Bank Policy Research Report.

Goldin, C. and Katz, L. (1997). *Why the United States led in education: lessons from secondary school expansions, 1910 to 1940* (Working paper series, No. 6144), NBER.

Golgher, A. (2010). *Diálogos com o Ensino Médio 2: O modelo Profluxo e outros indicadores demográficos* (Texto para discussão, No. 392). Belo Horizonte: Cedeplar/UFMG.

Hair, J., Black, W., Babin, B., Anderson, R. and Tatham, R. (2009). *Análise multivariada de dados*. Porto Alegre: Bookman.

Hanushek, E. and Wossmann, L. (2007). *The role of school improvement in economic development* (Working paper series, No. 12832). NBER.

Haussmann, S. and Golgher, A. (2016). Shrinking gender wage gaps in the Brazilian labor market: an application of the APC approach. *Nova Economia*, v. 26, p. 429-464

INEP (2008). Sinopse Estatística da Educação Básica. Accessed on 01 March 2014: http://www.inep.gov.br/basica/censo/Escolar/Sinopse/sinopse.asp.

Jain, A. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31 (8), 651-666.

LeSage, J and Pace, R. (2009). *Introduction to spatial econometrics.* Boca Raton : Taylor & Francis Group.

Lima, L. (2011). Da universalização do ensino fundamental ao desafio de democratizar o ensino médio em 2016: o que evidenciam as estatísticas? Revista Brasielira Estudos Pedagógicos, 92 (231), 268-284.

Meyer, J., Ramirez, F. and Soysal, Y. (2012). World expansion of mass education 1870-1980. *Sociology of Education*, 65 (2), 128-149.

Neri, M. et al. (2008). *Motivos da evasão escolar*. Accessed on 01 march 2014: http://www.fgv.br/cps/tpemotivos/

PISA (2009). *Results: what students know and can do: student performance in reading, mathematics and science*. Accessed on 01 march 2014: http://www.pisa.oecd.org/document/61/0,3746,en_32252351_32235731_46567613_1_1_1_1,00.html

Pontili, R. and Kassouf, A. (2008). Is age-grade distortion in Brazil´s primary public education system more closely associated to school infrastructure or to family characteristics? *Well-being and social policy*, 4 (1), 29-54.

Riani, J. and Golgher, A. (2004). *Indicadores educacionais confeccionados a partir de bases de dados do IBGE*. En: J. Riani and E. Rios-Neto (Eds.), Introdução à Demografia da Educação (pp. 89-128). Campinas: ABEP.

Rodrigues, C., Rios-Neto, E. and Pinto, C. (2013). Changes in test scores distribution for students of the fourth grade in Brazil: A relative distribution analysis for the years 1997-2005. *Economics of Education Review*, 34, 227-242.

Rodrigues, C., Rios-Neto, E. and Pinto, C. (2011). Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997 a 2005. *Revista Brasileira de Estudos de População*, 28 (5), 5-36.

Schwartzman, S. and Cossio, M. (2007). Juventude, educação e emprego no Brasil. *Cadernos Adenauer - Geração Futuro*, VII (2), 51-65.

Soares, T. (2005). Modelo de três níveis hierárquicos para a proficiência dos alunos de 4ª série avaliados no teste de língua portuguesa do SIMAVE/PROEB-2002. *Revista Brasileira de Educação*, 29, 73-87.

Soares, J. and Alves, M. (2013). Effects of schools and municipalities in the quality of basic education. *Cadernos de Pesquisa,* 43 (149), 492-517.

Tselios, V. (2008). Income and educational inequalities in the regions of the European Union: geographical spillovers under welfare state restrictions. *Papers in Regional Science*, 87 (3), 403-430.

UNESCO (2009). Education indicators – Technical guidelines. Accessed on 01 march 2014, Unesco Institute for statistics: http://www.uis.unesco.org/Library/Documents/eiguide09-en.pdf.

Urquiola, M. and Calderon, V. (2006). Apples and oranges: educational enrollment and attainment across countries in Latin America and the Caribbean. *International Journal of Educational Development*, 26, 572-590.