The Annals of Regional Science (2021) 66:255–277
https://doi.org/10.1007/s00168-020-01020-3

ORIGINAL PAPER

Spatial distribution and dissemination of education in Brazilian municipalities

Laura Desirée Vernier Fujita[1] · **Izete Pengo Bagolin**[1] · **Adelar Fochezatto**[1]

Received: 15 May 2019 / Accepted: 3 September 2020 / Published online: 22 September 2020

Abstract

This article aims to investigate the spatial effect of the quality of human capital and to identify if there is educational spillover from higher education in the proficiencies of elementary students. Although the role of geographical aspects has been studied in several academic areas, in the literature on the economics of education such studies are incipient and the few existing contributions use the stock of education as the object of analysis. The present study differs by using proficiency as a proxy for quality and also by examining the spillover effect of the presence and quality of university courses in the area of education. The results indicate a strong spatial dependence, suggesting the spatial dimension influences school performance. On the other hand, higher education institutions influence school performance in the municipality of origin, but not that of neighboring municipalities.

JEL Classification I21 · I23 · R19

1 Introduction

An effective educational system is among the main objectives of any government in pursuit of development and economic growth. Efforts to improve the education system are justified by the various benefits education provides in both the individual and social fields. Among the benefits for individuals are greater opportunity within the labor market and the capacity to generate future income (Murnane et al. 1995; Murphy and Peltzman 2004; Menezes-Filho 2001). In turn, society benefits from the country's economic growth and development, more specifically from the decreased

✉ Laura Desirée Vernier Fujita
laura.vernier@pucrs.br

Izete Pengo Bagolin
izete.bagolin@pucrs.br

Adelar Fochezatto
adelar@pucrs.br

[1] Economics at PUCRS, Porto Alegre, Brazil

likelihood of unemployment, reduced crime, improved health levels and increased productivity (Hanushek and Kimko 2000; Barros and Mendonça 1997; Bishop 1989). Although efforts by the Brazilian government have resulted in educational advances in terms of years of schooling, children's access to schools and a slight increase in student achievement (Soares et al. 2016), Brazilian students still score below the OECD average and the goals set in the National Education Plan (OECD 2019; Soares and Alves 2013).

Hence, studies using Brazilian data have sought to identify the main determinants of educational achievement, focusing mainly on school performance (Alves and Franco 2008; Alves and Soares 2013; De Andrade and Soares 2008; Brooke and Soares 2008; Franco et al. 2007; Reynolds and Teddlie 2008) and socioeconomic background (Alves and Soares 2009, 2013; Lee 2008). Studies have also analyzed the effects of per-student public expenditure on education (Amaral and Menezes-Filho 2008; Carreira and Pinto 2007; Diniz 2012; Fernandes 2013).

In general, the findings from the literature, in which various methodologies and databases are adopted, indicate that, when the characteristics of the student and family background are controlled, school characteristics have little impact on student achievement (Menezes-Filho 2007; Machado et al. 2008; Soares 2005).

The study by Vernier et al. (2015) shows that when school performance is analyzed in a more aggregated way, in addition to the socioeconomic issues of students, the characteristics of teachers, principals and the school also have an impact. Furthermore, the authors find that the region to which the student belongs also affects their school achievement, highlighting the importance of space.

Despite the decentralization of the Brazilian educational system that occurred in the 1990s (Da Mota 2008), with which most of the responsibility for primary education was attributed to municipalities, few studies take municipalities into account when studying student achievement. Among them, Ceneviva (2010) and Leme et al. (2009) measured the effect of municipal management on student achievement and Soares and Alves (2013) analyzed the effect of schools and municipalities on the inequality of the educational results.

Studies that investigate the effect of human capital in the neighborhood are commonly found in the literature. Such studies focus on assessing how performance and economic growth are influenced by the performance of nearby localities (Easterly and Levine 1995; Moreno and Trehan 1997; Ramos et al. 2010; Rosenthal and Strange 2008; Moretti, 2004; Beeson and Montgomer 1993). Case and Rosen (1993) provide evidence to show US government spending is positively affected by neighboring states. Spatial technological interdependence has also been explored by Ertur and Koch (2007).

Most of the Brazilian empirical studies related to the spatial effect of human capital have focused on the relation with economic growth, productivity and wages of the neighboring regions (Silveira Neto 2001; Magalhães et al. 2005). Research quality and innovation spillovers (Anselin et al. 1997) as well as the university–industry linkage (Garcia, Araujo and Mascarini 2013) have also been analyzed.

Notwithstanding the importance of human capital in the context of regional disparities and the significant amount of research that has sought to understand the determinants of school performance in Brazilian education, there has been little effort to investigate the distribution and spatial effect of educational quality. Hence, there is a lack of research regarding the effect of the quality of human capital in neighboring regions. Although the role of location has been explored from many perspectives, the role of spatial dependence on the performance of the Brazilian students requires further investigation.

Given the importance of location, and recognizing that proximity facilitates the exchange of information and knowledge, the present study seeks to contribute toward filling that gap in the literature. Taking the hypothesis that the direct impacts of HEI on local school achievements is unknown (we supposed to be positive, if any) and that the spatial scale of the impact is even more difficult to predict, this paper tries to specify these issues decomposing the effects in the local and neighboring municipalities. Thus, the main aim is to investigate if the quality of higher education has any effect on the school student achievement. For that sake, we use spatial econometric models, which permit the study of externalities, thus allowing us to identify if the variables used in the model of a given region have spillover effects in neighboring regions.

In this field, the noteworthy study by Ertur and Koch (2007) shows that the knowledge accumulated in a given country depends on the knowledge accumulated in neighboring countries. In estimating a model of technological interdependence, the authors found that the stock of knowledge in one country generates externalities that can cross national boundaries and affect other countries, and that this effect decreases with geographic distance. The authors attribute this spatial reflex to 'learning by doing.'

Given the importance of Ertur and Koch's study (2007) and considering that it focused on the stock of human capital, the present study seeks to contribute to the debate regarding the importance of the spatial dependence in educational quality by analyzing the proficiency scores of schools, a variable closely related to the quality of teaching.

In addition to testing the effect of the quality of HEI on the performance of schools in local and neighboring municipalities, it is also important to test the potential autocorrelation with the socioeconomic and cultural particularities of Brazilian municipalities. The inclusion of these aspects is important due to the heterogeneities of size, family background, gender composition for, example.

From the theoretical perspective, we adopted an educational production function which is used to identify the existing relationship between a series of inputs to the educational process and its product. We go further in comparison with previous studies by including the quality of higher education as an input. The importance of analyzing educational quality is highlighted in the studies by Hanushek and Kimko (2000), according to whom a good educational system does not only arise from the stock of years of study, but also from the quality of teaching.

Thus, the present study proposes to carry out an analysis of the spatiality of education with a focus on its quality. That is, it seeks to investigate if the spillover effect of human capital stock identified by Ertur and Koch (2007) is also found in the quality of education in Brazilian municipalities. In addition, it seeks to establish if there is educational spillover through higher education, through better qualified teachers. In other words, whether higher education influences school education in the respective municipality and neighboring municipalities.

The study has 6 sections. Following this introduction: Sect. 2 presents the data used and the descriptive statistics; Sect. 3 describes the methodology; Sect. 4 section details the methods adopted in the spatial matrix analysis; Sect. 5 discusses the results found; and finally, Sect. 6 contains the closing remarks.

2 Data and descriptive statistics

To achieve the proposed objective, information on 5507 Brazilian municipalities, covering the period between 2008 and 2013, was collected from two data bases. The first is The Basic Education Assessment System (Portuguese—*Sistema de Avaliação da Educação Básica*—SAEB), which is an assessment of Brazilian schools carried out by INEP. The exam tests student achievement in Mathematics and Portuguese. In addition to the tests, the SAEB includes questionnaires that are administered to principals, teachers and students, allowing for a more comprehensive assessment of the schools. The second is the National Examination of Student Performance (Portuguese—*Exame Nacional de Desempenho dos Estudantes*—ENADE), which is also conducted by INEP, but assesses courses in higher education institutions (HEIs) in Brazil. The examination was proposed in 2003 and formally instituted in 2004. The set of courses is divided into three blocks, and only one of these blocks is assessed each year. Based on the student achievement, the test, totaling 40 questions, is divided into two parts: the first on general knowledge (GK) and the second on a specific component (SC). In the GK component, integral elements of the professional profile are considered, such as ethical attitude, social commitment, capacity for critical analysis. The SC assesses issues specific to the area of knowledge of the course.

The use of higher education data in the study is based the assumption that the best performing HEIs produce better qualified professionals, including in the area of education and thus would be associated with better grades in elementary school education. The impact of university quality on improvement in local human capital and its regional spillover was explored by Beeson and Montgomer (1993) using US data. For the Brazilian context, this assumption is important because the universities have an important role in supporting the National Plan for Training of Basic Education Teachers, developed by the Coordination for the Improvement of Higher Education Personnel (CAPES). This plan is designed to mitigate the critical problems identified in the teacher training process in the country's elementary education system (Castro et al. 2018).

This paper investigates the effect HEIs have on school student achievement in the municipality and neighboring municipalities. The dependent variable is the municipal average score in mathematics proficiency in 2013, obtained in the SAEB. We chose this discipline because mathematics is considered a universal language, which permits comparisons with international tests and studies. In addition, the interpretation of mathematics problems requires a certain level of language skills.

According to Silva and Hasenbalg (2002), the effect of family background tends to be lower as from the second half of elementary school. Thus, the proficiency

Table 1 Number of pupils by administrative level. Source: Elaborated by the authors (SAEB 2013)

Administrative level	Number of Pupils	% in relation to the total (%)
State	1,635,535	60.1
Municipal	1,023,065	37.6
Private	56,981	2.1
Federal	5,007	0.2
Total	2,720,588	100.0

scores used in this study will be those of 8th and 9th grade students, since they are more susceptible to public policy measures.

Table 1 shows the number of 8th and 9th grade students who took the exam in 2013 according to the school administrative level. There is a notable predominance of state and municipal schools, which account for 60.12% and 37.6% of the sample, respectively.

The SAEB proficiency scores range from 0 to 500, and in 2013 among the municipalities the average score in mathematics was 244.9 points in the 8th and 9th grades, while the maximum score was 320.72. Figure 1 shows the distribution of school education in the Brazilian municipalities. It can be seen that most of the highest scoring municipalities are concentrated in the Centre-West, Southeast and South regions.

To determine the effect of higher education on school education, a time lag is necessary, since the time needed to complete the course and to start working as a teacher must be considered. Thus, the data collected on the performance of HEIs will refer to the year 2008.[1] The key variable for higher education will be the ENADE scores for the degree courses in Language and Literature, Mathematics and Education of the various institutions. In view of the fact that the ENADE score is a measure of the effect of the quality of higher education, and that in the absence of an HEI, it is impossible to identify this effect, municipalities that do not have an HEI within their territories will be attributed a score of zero, for estimation purposes.

Of the 5507 municipalities, 739 had HEIs offering degree courses in language and literature, mathematics and education in 2008. Figure 2 shows the distribution of those institutions. The ENADE score varies from 0 to 5, while the average reached by the municipalities was 2.86.

To capture the effect of higher education, it is necessary to control for the effect of the other variables that affect school performance. To do so, variables consolidated in the education literature, such as parental schooling and other socioeconomic characteristics, are used. Table 2 shows the dependent and explanatory variables used on this paper.

[1] The data available from the ENADE for the degree courses in Language and Literature, Mathematics and Education refer to the years of 2005, 2008 and 2011. Due to the lack of municipal code information in the year 2005, the data referring to 2008 were used.

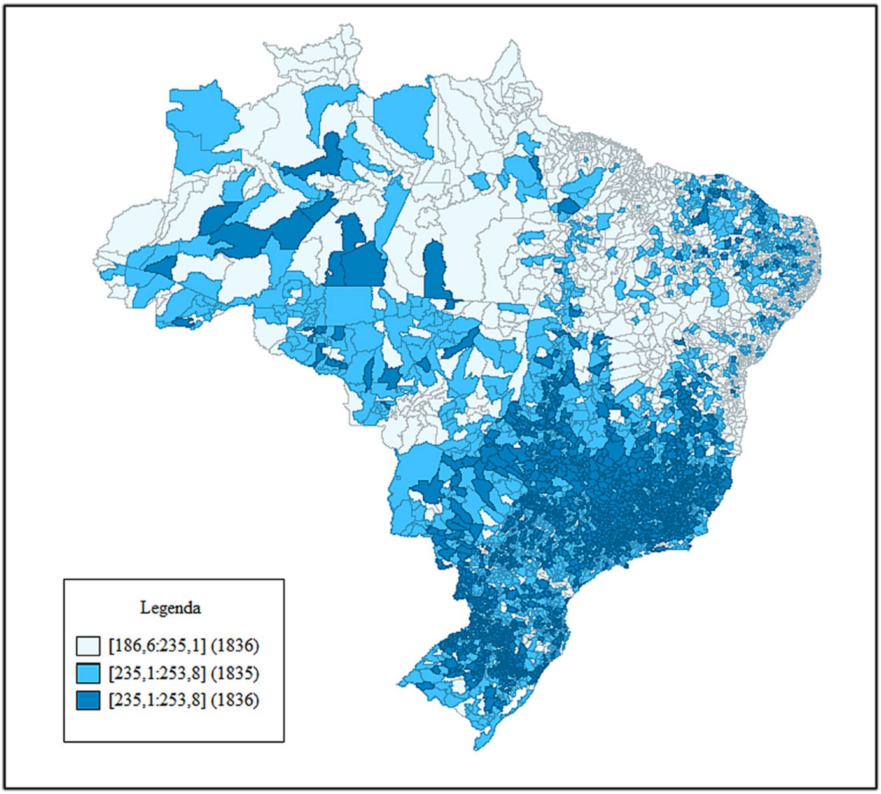

Fig. 1 Distribution of School Education by Brazilian Municipality

Parental schooling is introduced in the model as the proportions of mothers and fathers that have completed high school. The data for this variable are obtained from the students' questionnaire within the SAEB, through the question: "Up to what grade did your mother or the woman responsible for you study? Up to what grade did your father or the man responsible for you study?" "Appendix 1" presents possible alternative responses. In order to consider parents who have completed high school, alternatives showing parents educated up to high school will be the focus of this variable.

The *mother's and father's schooling* are standard variables used in education production function to test the impact of family background. The assumption being that the higher the parent's educational level, the higher the student's achievement. On average, in Brazilian municipalities 33% of mothers and 26% of fathers had completed high school in 2013. Based on the microdata, "Appendix 2" presents the percentage of fathers and mothers in each level of schooling according to the answers given by the pupils.

Student's *sex* and *color* are also included in the model due to their importance in determining performance (Soares 2005). The evidence in the literature shows

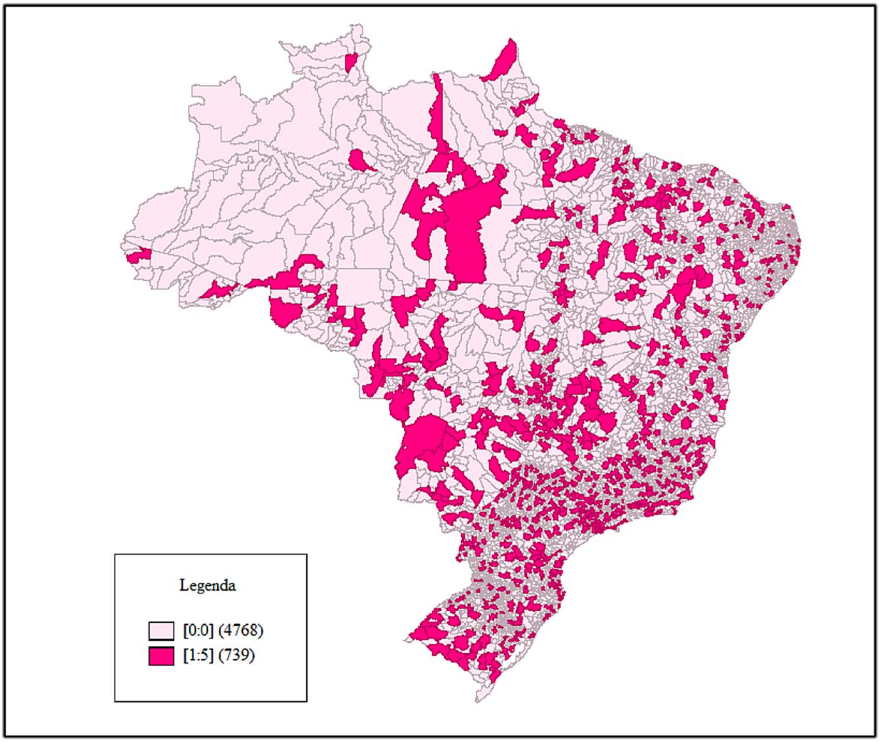

Fig. 2 Distribution of Higher Education by Brazilian Municipalities

males, generally, perform better than females in math (Le and Nguyen 2018; Guiso et al. 2008). According to Soares (2003), white Brazilian students tend to perform better at school. Along the same lines, the studies carried by Albernaz et al. (2002), Soares (2003, 2005) and França and de Gonçalves (2012) found that being black has a negative impact on school achievement. Thus, we expect white students to achieve higher educational outcomes.

Through the teachers' questionnaire, the SAEB provides information on teacher training. The item that addresses *post-graduate training* offers the following alternative responses: (a) Have you not done or not completed post-graduate training; (b) updated training; (c) specialization; (d) master's; (e) PhD. To obtain the proportion of this variable, a dummy was initially created (1: did post-graduate training, 0: otherwise) and, from this, the proportion of teachers with post-graduate education was calculated. Following the same hypothesis tested by Albernaz et al. (2002) and França and de Gonçalves (2012), we assume that having better qualified teachers will positively impact the student outcomes.

Teachers are considered hardworking if they assign and correct homework. Based on the SAEB questionnaire, this can be assessed based on the answers given to the following question: "Does the teacher correct Mathematics

Table 2 Variable description. Source: Elaborated by the authors

Variable	Name	Description	Source/Period
School Performance (dependent)	*eduS*	Municipal average performance in Mathematics Continuous variable—ranges from 0 to 500	SAEB 2013
Higher Education Performance	*eduH*	Municipal average performance in ENADE Continuous variable—ranges from 0 to 5	ENADE 2008
Mothers' Schooling	*SchMth*	Proportion of Mothers with High School Continuous variable—ranges from 0 to 100	SAEB 2013
Fathers' Schooling	*SchFth*	Proportion of Fathers with High School Continuous variable—ranges from 0 to 100	SAEB 2013
Sex	*Sex*	Proportion of Male Pupils Continuous variable—ranges from 0 to 100	SAEB 2013
Skin Color	*Col*	Proportion of White Pupils Continuous variable—ranges from 0 to 100	SAEB 2013
Post-Graduate Training	*PostGrad*	Proportion of Teachers with Post-Graduate training Continuous variable—ranges from 0 to 100	SAEB 2013
Teacher Effort	*TchEff*	Proportion of Hardworking Teachers Continuous variable—ranges from 0 to 100	SAEB 2013
Parental Stimulus	*Stimulus*	Proportion of Parents that Encourage their Children Continuous variable—ranges from 0 to 100	SAEB 2013

homework?" If the student answered 'always' or 'almost always,' the teacher is considered to be hardworking (1: hardworking teacher, 0: otherwise), and, from this, the proportion of hardworking teachers per municipality is calculated. As with teacher qualification, teacher engagement with the students' learning process will increase the students' test scores, so we assume positive results related to *teacher effort*.

Parental *Stimulus* has been shown to be an important factor for student success. This finding has been associated with a diverse range of desirable outcomes ranging from better test scores, higher expectations regarding the future, more confidence and better behavior at school (Singh 2015; Bashir and Bashir 2016). In this study, the variable *parental stimulus* is determined based on the questions and answers in "Appendix 3". If the answers to all the questions were 'always/almost always' or 'yes,' the parents are considered to encourage their children to study (1: parents encourage; 0: otherwise). From this dummy, the proportion of parents who

Table 3 Descriptive statistics for the variables used Source: Elaborated by the authors

Variable	Average	S.D.	Minimum	Maximum
eduS	244.9	19.84	186.59	320.72
eduH	2.86	1.02	0.00	5.00
SchMth	0.33	0.12	0.00	0.91
SchFth	0.26	0.12	0.00	0.82
Sex	0.47	0.07	0.00	1.00
Col	0.36	0.21	0.00	1.00
PostGrad	0.66	0.25	0.00	1.00
TchEff	0.65	0.11	0.04	0.97
Stimulus	0.39	0.10	0.00	0.77

encourage their children in the study is obtained. Table 3 summarizes the descriptive statistics for the variables used in the study.

In the next section, the econometric instruments used to identify if there is a spatial relation in education in the Brazilian municipalities will be presented.

3 Methodology

When using a spatial regression model, it is important to carry out a prior investigation regarding the existence of spatial dependence. Exploratory spatial data analysis (ESDA) allows us to evaluate if there is any pattern of spatial association between regions.

Implementing ESDA requires the construction of a spatial weight matrix (W), also known as the neighborhood matrix. The neighborhood can be attributed in several ways. Essentially, there are four matrices, namely: Rook (has at least one side in common), Queen (has at least one point in common), Distance (is a certain distance from the centroid) and K (determined number of the closest neighbors).

Spatial dependence arises when the value of the variable in a given location depends on the value of that variable in neighboring regions, that is, whether or not the data are randomly distributed in space. Moran's I is a method that allows us to test and infer spatial dependence.

In this study, initially the statistics that permit the identification of spatial patterns will be used: Moran's I global and local. The global Moran I provides a summary of the data's spatial distribution, that is, a single (average) value for all regions. The local Moran I (LISA), when calculating a value for each observed unit, allows the identification of different patterns of spatial distribution (clusters or outliers). The global Moran I can be defined as:

$$I = \left(\frac{n}{S_o}\right)\left(\frac{Z_t' W Z_t}{Z_t' Z_t}\right) \tag{1}$$

where Z_t is the vector of n regions for the year t in the form of deviation from the mean. W is the spatial weight matrix, and the term S_o is a scalar equal to the sum of all elements in W. The value of the index varies between -1 and 1, where a figures closer to -1 indicate a negative autocorrelation, and those closer to 1 indicate a positive autocorrelation.

Checking the local patterns and determining the regions that contribute most to the spatial autocorrelation can be done with the aid of LISA. This indicator was initially suggested by Anselin (1995) and can be calculated as follows:

$$I_l = \frac{y_j \sum_{j=1}^{n} w_{ij} y_i}{\sum_{i=1}^{n} y_i^2} \tag{2}$$

where n denotes the number of regions, w_{ij} are the elements of the spatial weight matrix, y_i and y_j are the values of the variable used, while i and j refer to the different localities.

In addition, we used Moran's Scatterplot, which establishes a graphical comparison of the values of the variable of the regions with the respective values resulting from the weights of the neighbors, which permits observation of any regional concentration. This diagram is divided into four quadrants, two of which correspond to the positive spatial correlation (clusters), which may be associated with two different patterns, High–High (HH) or Low–Low (LL). The former (located in the upper right) indicates regions in which the score for the variable of interest is above average which are surrounded by neighbors whose scores are also above average. The LL standard, on the other hand, indicates regions with low scores for the variable of interest, surrounded by regions that also have low scores.

The second and fourth quadrants indicate regions with negative autocorrelations (outliers) and can be represented by either the High–Low (HL) or Low–High (LH) standards. The HL (LH) quadrant indicates locations in which the score for the variable of interest is above (below) the average, while its neighbors have scores for the same variable below (above) the average.

Once the spatial heterogeneity is measured, the next step is to include spatial dependence in the model to be studied. For this, a spatial econometric model is used. Unlike conventional models, spatial econometrics makes it possible to consider the effects of spatial dependence. The dependency is incorporated into the model through a matrix of spatial weights, capturing the data generation process. As mentioned earlier, there are several types of weight matrices. To identify which one best captures the spatial structure of the data, Moran's I is calculated by choosing the one that produces the most value (greatest spatial autocorrelation). In this study, the matrix K closest neighbors ($K=1$) was chosen. The four models estimated below use this matrix.

Generally, the starting point for econometric analysis models is the classical linear regression model, estimated by ordinary least squares (OLS). However, Ertur and Koch (2007) suggest spatial dependence or heterogeneity may lead to unreliable OLS estimates due to the possibility of heteroscedasticity generated by changes in coefficients or error variance between observations.

To better explore the spatial nature of the problem proposed in the present study, the hypothesis to be tested is described by the following general equation:

$$\begin{aligned} y_i &= \alpha + \beta X_i + \rho W_{ij} y_j + W X_i \theta + \delta_i \\ \delta_i &= \lambda W_{ij} \delta_j + \varepsilon_i \\ \varepsilon &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{3}$$

where y_i (y_j) is the variable explained in the region i (j), α is the intercept, X is the matrix nxk of explanatory variables, β is the vector k×1 of coefficient, ρ is the parameter related to the spatial lag of the explained variable, λ is the noise variance parameter, θ is the spatial autocorrelation coefficient related to the explanatory variables, W is the spatial weight matrix n×n, with $W_{ij} > 0$, when the region j is close to the region i.

By changing the values of the parameters ρ and λ, different models are obtained. In the case where there is no spatial dependence, neither in the dependent variable nor in the disturbances ($\rho=0$, $\lambda=0$ and $\theta=0$), the model would be the traditional OLS, which can be represented as follows:

$$y_i = \alpha + \beta X_i + \varepsilon_i \tag{4}$$

In the case of $\rho \neq 0$ ($\lambda=0$ and $\theta=0$), the spatial autoregressive (SAR) model is estimated. Thus, spatial dependence is included in the model by spatially lagged values of the dependent variable, as described in the following equation:

$$y_i = \alpha + \beta X_i + \rho W_{ij} y_j + \varepsilon_i \tag{5}$$

The spatial error model (SEM), in turn, reflects the spatial dependence on the residuals (ρ = 0, θ = 0 and $\lambda \neq 0$). This specification indicates that a random shock introduced in one region affects the others through the spatial structure. The model can be specified as follows:

$$\begin{aligned} y_i &= \alpha + \beta X_i + \delta_i \\ \delta_i &= \lambda W_{ij} \delta_j + \varepsilon_i \\ \varepsilon &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{6}$$

The spatial Durbin model (SDM), in turn, reflects the spatial dependence on the endogenous variable and on the explanatory variables ($\rho \neq 0$ and $\lambda \neq 0$). The model can be specified as follows:

$$\begin{aligned} y_i &= \alpha + \beta X_i + \rho W_{ij} y_j + W X_i \theta + \varepsilon_i \\ \varepsilon &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{7}$$

According to Anselin and Bera (1998), the most appropriate estimation for these models is either that of maximum likelihood or of instrumental variables, since OLS estimates generate biased and inconsistent results due to the simultaneity in the nature of the autocorrelation caused by the spatial lag. The next section will describe the method for determining the spatial weights matrix and the matrix that will be used in the study.

4 Determining the weights matrix

To identify the spatial interactions between municipalities, it is necessary to specify how these regions are connected. The chosen matrix is in accordance with the structure of the sample. While the use of the distance matrix allows the identification of the proximity of municipalities, it does not allow us to see if there is a border between them. And, since Brazilian municipalities are not homogeneous in terms of size, the use of a weighting matrix based on distance or contiguity might generate an unbalanced structure. A common solution to this problem is to consider the weight matrices based on the nearest neighbors, since this would force each unit to have the same number of neighbors (Anselin 2002; De Dominicis et al. 2013).

In order to specify the number of neighbors that will compose the neighborhood matrix, Almeida's criterion (2012) is used. According to this criterion, after testing the Moran's I for a set of matrices, the matrix that generated the highest value, and that is statistically significant is selected. The Moran's I coefficient is presented in "Appendix 4."

In this study, the spatial weights matrix is used with the closest neighboring municipality ($k=1$). With the definition of the matrix and with the aid of the local Moran I, it is possible to study the local spatial patterns in more detail. The following figures present the LISA for school and higher education performance according to the selected matrix.

Figure 3 shows the presence of high-performance clusters in the Midwest, Southeast and South. That is, counties with above-average performance are surrounded by neighbors whose values are also above average. On the other hand, in the northern and northeastern regions, there are low performance clusters.

While there is a positive spatial correlation for school education performance, in higher education there is a negative spatial correlation (outliers). These are mostly represented by high–low standards, that is, municipalities with high-performance higher education surrounded by poorly performing neighbors (Fig. 4).

Having defined the models and the spatial weights matrix, it is possible to estimate the spatial models. The following sections present the results found and the final remarks regarding the present study.

5 Results and discussion

The results shown were obtained through the following steps. The first step was to estimate the model using ordinary least squares (OLS), in order to detect the occurrence of spatial autocorrelation and test the most appropriate spatial model. Thus, four regressions were estimated with the variables previously presented. The following equations present the specification of the models:

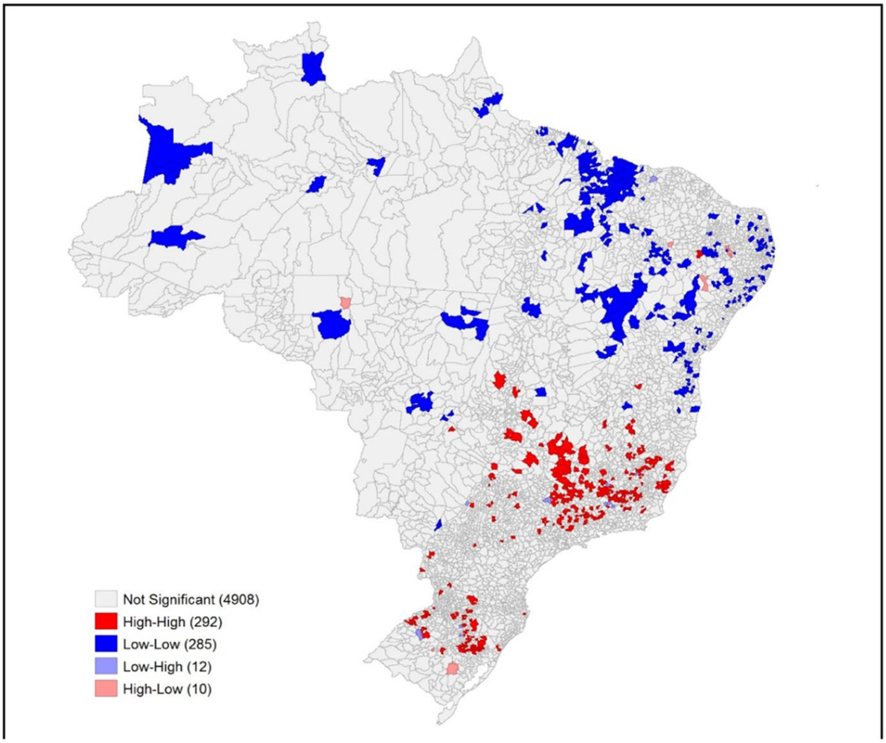

Fig. 3 LISA for School Educational Performance

$$eduS_i = \beta_0 + \beta_1 eduH_i + \beta_2 Col_i + \beta_3 Sex_i + \beta_4 SchMth_i + \beta_5 SchFth_i + \beta_6 PostGrad_i \quad \text{(Model 1)}$$

$$eduS_i = \beta_0 + \beta_1 eduH_i + \beta_2 Col_i + \beta_3 Sex_i + \beta_4 SchMth_i + \beta_5 SchFth_i + \beta_7 TchEff_i \quad \text{(Model 2)}$$

$$eduS_i = \beta_0 + \beta_1 eduH_i + \beta_2 Col_i + \beta_3 Sex_i + \beta_6 PostGrad_i + \beta_8 Stimulus_i \quad \text{(Model 3)}$$

$$eduS_i = \beta_0 + \beta_1 eduH_i + \beta_2 Col_i + \beta_3 Sex_i + \beta_7 TchEff_i + \beta_8 Stimulus_i \quad \text{(Model 4)}$$

Model 1 is the basic model. In model 2, the aim is to check if *teacher effort* has the same effect as the *teacher post-graduate training*. In model 3, we want to see if the effect of the '*parental stimulus'* is in line with the '*mother's and father's schooling.'* Finally, model 4 simultaneously checks whether the *teacher effort* and *parental stimulus* are aligned with teacher *post-graduate training* and *mother's and father's schooling*.

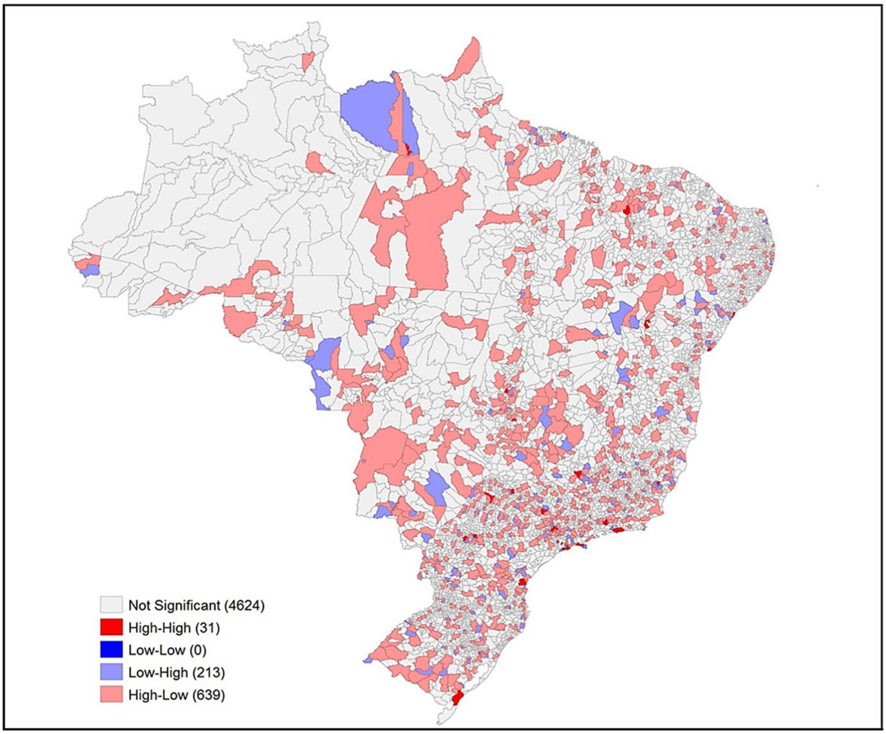

Fig. 4 LISA for Higher Education Performance

Table 4 Lagrange multiplier tests *p*-values. Source: Elaborated by the authors

Test	Model	Model 1	Model 2	Model 3	Model 4
MLρ (Lag)	SAR	1315.8***	1259.9***	1381.2***	1268.9***
ML*ρ (Robust Lag)	SAR	25.6***	131.7***	32.7***	124.4***
MLλ (Error)	SEM	1310.1***	1140.2***	1377.0***	1150.6***
ML*λ (Robust Error)	SEM	19.9***	11.9***	28.6***	6.1**
MLρλ (Lag and Error)	SARMA	1335.7***	1271.8***	1409.8***	1274.9***

*Denote *p*-values less than 10%

**Denote *p*-values less than 5%

***Denote *p*-values less than 1%

The OLS was estimated for each model (the results can be seen in "Appendix 5"), after which the Lagrange multiplier test was used to see whether there was an effect of the spatial structure on the model. The Lagrange multiplier test values for the four estimated models are shown (Table 4):

The last four columns show the results found for each model. The first two lines indicate whether the dependent variable has a spatial effect. The third and fourth lines present the results for the spatial relation of the residuals. Finally,

Table 5 Results found with models 1 to 4, using spatial Durbin modeling (SDM). Source: Elaborated by the authors

Variables	Model 1	Model 2	Model 3	Model 4
Const	124.130***	96.098***	111.684***	100.229***
eduH	0.450**	0.774***	1.428***	1.505***
Col	29.168***	25.546***	29.357***	27.408***
Sex	− 0.438	2.508	4.235	4.993**
SchMth	17.551***	12.138***		
SchFth	7.677***	15.999***		
PostGrad	1.642**		1.245*	
TchEff		54.833***		51.203***
Stimulus			41.001***	3.107
W_eduH	−0.298	−0.150	−0.203	−0.087
W_Col	−2.795	−2.086	−2.616	−2.951
W_Sex	8.934***	7.268***	10.892***	10.874***
W_SchMth	−7.312***	− 10.833***		
W_SchFth	−0.812	5.538**		
W_PostGrad	−0.449		− 1.149	
W_TchEff		−4.225**		9.588***
W_Stimulus			− 17.155***	−25.456***
ρ(rho)	0.413***	0.395***	0.431***	0.402***
Log likelihood (LIK)	−22,546.18	−22,066.27	−22,412.00	−22,156.47
Akaike info criterion (AIC)	45,122	44,163	44,850	44,339
R^2	0.523	0.594	0.550	0.583
N	5507	5507	5507	5507

*Denote *p*-values less than 10%

**Denote *p*-values less than 5%

***Denote *p*-values less than 1%

the last line tests whether the residuals and the dependent variable have a spatial effect.

The estimation tests showed statistical significance for spatial dependence, indicating the spatial structure impacts the quality of education. These results demonstrate both the dependent variable, and the residuals have a spatial effect. In such conditions, the literature suggests the use of spatial Durbin models (SDMs), in which the estimators are unbiased and the statistical tests are valid even with spatial autocorrelation in the residual (Elhorst 2010; Loonis and de Bellefon 2018). To this end, four SDMs were estimated. Table 5 shows the estimated results, and Table 6 shows the direct and indirect effects.

The models seek to identify the relationship between the quality of higher education and the quality of school education and the educational spillovers. The first model estimates the effect of teachers having a post-graduate training. However, this training effect may also, to some extent, capture the teacher pedagogy and teacher

Table 6 Direct, indirect and total estimated effects for the spatial Durbin models (SDM) Source: Elaborated by the authors

Models	Variables	Direct	Indirect	Total
(1)	*eduH*	0.416	−0.156	0.260
	Col	32.090	12.912	45.002
	Sex	2.317	12.180	14.497
	SchMth	17.537	−0.065	17.472
	SchFth	8.422	3.291	11.713
	PostGrad	1.715	0.321	2.036
(2)	*eduH*	0.820	0.211	1.031
	Col	27.916	10.879	38.795
	Sex	4.951	11.216	16.168
	SchMth	10.353	−8.195	2.158
	SchFth	19.509	16.109	35.619
	TchEff	59.995	23.696	83.692
(3)	*eduH*	1.566	0.587	2.153
	Col	32.702	14.303	47.005
	Sex	8.472	18.119	26.590
	Stimulus	41.175	0.741	41.917
	PostGrad	1.041	−0.872	0.169
(4)	*eduH*	1.662	0.709	2.372
	Col	29.849	11.050	40.899
	Sex	8.891	17.645	26.536
	Stimulus	−4.216	− 33.157	− 37.373
	TchEff	60.332	41.328	101.660

commitment to the students learning process. To filter this effect, we estimated model 2, including the variable *Teacher Effort (TchEff)*. However, with the inclusion of that variable, the multicollinearity problem reappeared. Thus, the model was tested without the variable *post-graduate training*, thus solving the problem. The association between the two variables may indicate that dedicated teachers seek better training.

Models 3 and 4 include the variable *parental stimulus* and, seeking to avoid multicollinearity, the variables *mother's schooling* and *father's schooling* are removed.

Since the SDM estimator expresses the spatial effect of the all the variables, including the explained variable, estimating the four models allows us to identify the spillover effect of Higher education on the neighboring schools.

Table 6 shows the estimated results for the direct, indirect and total effects. The direct effect, for example, illustrates how a shock in the explanatory variables affects the explained variable in the same municipality and the indirect effect measures the effect in the neighbors, which may in turn reflect that impact back on the municipality of origin. Thus, the indirect effect is obtained by subtracting the total effect from the direct effect. It is also possible to conclude that the direct effects are similar to the non-spatial models' coefficients.

The quality of education effect is expressed in all the estimated models through the Higher Education (*eduH*) and the spatial dependence of the explained variable

(ρ). The positive and significant coefficient of *eduH* indicates that municipalities with better Higher Education performance are more likely to have better achieving schools. This result is in accordance with the hypothesis that the best performing HEIs produce better qualified professionals, and this is associated with better grades in elementary school education.

In addition to the effect of higher education on school grades, the spatial dependence of the explained variable (ρ) is captured, suggesting school performance is positively associated with school performance in neighboring municipalities. The significance of the spatial parameter indicates the presence of education spillover among Brazilian municipalities.

Of the variables related to the characteristics of the students, *skin color* (*Col*) was significant and positive. The proportion of male students (*Sex*) only presents significant and positive results in the fourth model. With respect to parental schooling (*SchMth* and *SchFth*), both variables are positive and significant in models 1 and 2. There is a notable difference in relation to the magnitude of the coefficients, in which the values of the mother's schooling are higher than those of the father.

In model 3, *Parental stimulus* (*Stimulus*) was positive and significant, as expected. According to this information, the municipalities with the highest proportion of parents who encourage their children to study are associated with better school grades. However, when we introduce *teacher effort*, in model four, the significance of *parental stimulus* disappears. This may be because parents who are more concerned about educational importance seek better schools with more committed teachers.

In models 1 and 3, *PostGrad* is positive and statistically significant, indicating the higher the proportion of teachers with post-graduate qualifications there are, the better the students perform at school. *Teacher effort* (*TchEff*) and commitment to the students, expressed in models 2 and 4, are also positive and statistically significant.

Of particular interest in our analysis are the variables Higher Education Performance, *Post-Graduate Training* and *Teacher Effort*. Regarding the effect of Higher Education Performance, as the coefficient associated with *W_eduH* was found to have no statistical significance, higher education cannot be said to have a spatial spillover effect on elementary schools in neighboring municipalities. Therefore, the presence of a reputable university in a municipality has a local effect, increasing the local school performance.

The variable *W_PostGrad* is statistically insignificant. Thus, at the same time as there is a positive relation between the percentage of teachers with post-graduate qualifications and school performance in the municipality itself, there is no such relation with neighboring municipalities. This may arise due to a process of concentration, where teachers with better qualification are attracted to better job opportunities in municipalities with high-performing schools.

Lastly, in model four, the coefficient associated with *W_TchEff* is positive and significant, as expected, indicating that in addition to the direct positive effect, teacher effort also has a positive indirect effect, which means there is spatial spillover.

Considering these results and using information from the descriptive statistics (Table 3), it can be said that, in relative terms, there is a higher spatial concentration of teachers with graduate degrees (greater standard deviation) and that their knowledge does not overspill to schools in neighboring municipalities. On the other hand, the efforts made by teachers are relatively less concentrated and help improve the performance of students in the municipality of origin as well as that of students in neighboring municipalities.

In summary, in this study our aim was to verify the spatial spillover of the quality of higher education (*W_eduH*) in the mathematics performance of students in elementary schools (*edu_S*). This variable was not statistically significant. In relation to the lags of the other explanatory variables, it is difficult to establish a theoretical relationship. The negative signs found for some of the variables may be related to factors such as social inequality and distance. In Brazil, more developed municipalities are often surrounded by poor municipalities, which may explain why the same variable has contrasting effects in neighboring municipalities. Additionally, the distance between municipalities is very heterogeneous, which may cause the variable to overflow in the regions where the municipalities are closest and not where they are most distant.

6 Concluding remarks

The present study has sought to contribute to the literature on determinants of school performance through the use of spatial models that allow us to identify the existence of spatial dependence in relation to the quality of education. More specifically, whether the quality of education in a Brazilian municipality is associated with the education of neighboring municipalities.

The study identified a strong spatial dependence, indicating that spatial structure influences the quality of education. There was an educational spillover among Brazilian municipalities at the school level. In other words, one municipality's school performance is positively associated with the school performance of its neighboring municipalities.

Our main hypothesis that the best performing HEIs produce better qualified teachers, and thus would be associated with better grades in school, was only confirmed at the local level. The main contribution of the paper was to show that municipalities with better HEIs have better performing schools.

However, contrary to our expectations this positive local effect does not cross municipal boundaries. While being contrary to our hypothesis, this is an interesting result for consideration in further new studies and public policies decisions. The absence of a spatial effect of HEIs may be due to the concentration of universities in a small number of densely populated municipalities (739 or 13.4%) while schools are distributed throughout all the 5507 municipalities. Being a developing country with high levels of inequality, Brazil still has teachers who do not hold higher education degrees, despite an attempt to engage the universities in teacher training. As mentioned above, since 2009, the National Plan for Training of Basic Education Teachers has provided special classes in higher education courses, exclusively for public school educators who do not have higher education degrees in the area in which they work.

Overall, our findings highlight the importance of university–school integration and provide useful insights for consideration when deciding on public policy.

In Brazil, a large proportion of universities are public or community (federal, state or municipal) and they play an important role in regional development. Their contribution to human capital, technology, university–industry linkages has been well documented in the Brazilian literature (Silveira Neto 2001; Magalhães et al. 2005; Anselin et al. 1997; Garcia et al. 2013). Importantly, our study is the first to explore their local and spatial impact on the student outcomes.

Funding We thank the funding provided by the National Council for Scientific and Technological Development -CNPq – 306707/2017-9.

Appendix 1—responses regarding parent schooling

Response	Specification
A	Never studied
B	Did not complete the 4th grade (formerly primary)
C	Completed 4th grade, but did not complete 8th grade (formerly gymnasium)
D	Completed 8th grade, but did not complete High School (formerly 2nd grade)
E	Completed high school, but did not complete college.
F	Completed the College
G	Do not know

Source: Elaborated by the authors based on the SAEB

Appendix 2—percentage of mothers and fathers by level of schooling in 2013

	Mother (%)	Father (%)
Never studied	3.04	5.6
Never completed 4th grade	16.93	18.62
Completed 4th grade, but never completed 8th grade	22.06	20.17
Completed 8th grade, but never completed High School	17.13	19.29
Completed High School, but never completed a University degree	29.97	27.11
Completed University	10.88	9.2

Source: Elaborated by the authors

Appendix 3—questions about *Parental Stimulus*

Question	Answer
How often do your parents attend the parents meetings?	Always/ almost always
Do your parents encourage you to study?	Yes
Do your parents encourage you to do homework and assignments?	Yes
Do your parents encourage you to read?	Yes
Do your parents encourage you to attend school and not miss classes?	Yes

Source: Elaborated by the authors (SAEB, 2013)

Appendix 4—results of the variation of weight matrices

Matrix	Moran's I coefficient (*eduS*)
Queen (first order)	0.655
Rook (first order)	0.656
K (1 neighbor)	0.692
K (2 neighbors)	0.688
K (3 neighbors)	0.675
K (4 neighbors)	0.668
K (5 neighbors)	0.660
d (50 km)	0.634
d (100 km)	0.598
d (200 km)	0.543
d (300 km)	0.512
d (370 km)*	0.495

Source: Elaborated by the authors

K = matrices of closest K neighbors (inverse-distance); d = distance matrices (inverse-distance); all the matrices are row-standardized; (*) 370 km is the minimum distance to ensure no municipality is isolated

Appendix 5—results of the OLS models

Variables	Model 1	Model 2	Model 3	Model 4
Const	216.537***	173.263***	202.976***	181.549***
eduH	0.392*	0.804***	1.684***	1.770***
Col	44.893***	40.357***	47.082***	42.977***
Sex	5.915*	9.650***	13.598***	14.133***
SchMth	15.315***	7.436***		
SchFth	13.011***	26.144***		
PostGrad	1.449		0.415	
TchEff		66.524***		71.455***
Stimulus			40.329***	− 12.827***
R^2	0.280	0.400	0.295	0.368
N	5507	5507	5507	5507

Source: Elaborated by the authors

References

Albernaz Â, Ferreira FHG, Franco C (2002) Qualidade e equidade no ensino fundamental brasileiro. Pesquisa e Planejamento Econômico 32(3):453–476
Almeida E (2012) Econometria espacial aplicada. Alínea, Campinas
Alves MTG, Franco C (2008) A Pesquisa em eficácia escolar no Brasil: evidência sobre o efeito das escolas e fatores associados à eficácia escolar. In: Brooke N, Soares JF (eds) Pesquisa em eficácia escolar: Origem e trajetórias. Editora UFMG, Belo Horizonte, pp 482–500
Alves MTG, Soares JF (2009) Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. Opinião Pública 15(1):1–30
Alves MTG, Soares JF (2013) Contexto escolar e indicadores educacionais: condições desiguais para a efetivação de uma política de avaliação educacional. Educação e Pesquisa 39:177–194. https://doi.org/10.1590/S151797022013000100012
Anselin L (1995) Local indicators of spatial association—LISA. Geogr Anal 27(2):93–115
Anselin L (2002) Under the hood issues in the specification and interpretation of spatial regression models. Agric Econo 27(3):247–267
Anselin L, Varga A, Acs Z (1997) Local geographic spillovers between university research and high technology innovations. J Urban Econ 42(3):422–448
Anselin L, Bera AK (1998) Introduction to spatial econometrics. In: Handbook of applied economic statistics vol 237
Barros RP, Mendonça R (1997) Investimentos em Educação e Desenvolvimento Econômico. Rio de Janeiro: IPEA, (Texto para Discussão No. 525)
Bashir L, Bashir H (2016) A Study on Parental Encouragement among Adolescents. Education 5(4):269–270
Bishop J (1989) Is the test score decline responsible for the productivity growth decline? Am Econ Rev 79(1):178–197
Brooke N, Soares JF (2008) Comentário: seção 3: O que faz a diferença? Os métodos e evidencias da pesquisa sobre o efeito escola. In: Brooke N, Soares JF (eds) Pesquisa em eficácia escolar: origem e trajetórias. Belo Horizonte, Editora UFMG, pp 218–224
Carreira D, Pinto JMR (2007) Custo aluno-qualidade inicial: rumo a educação pública de qualidade no Brasil. Global, São Paulo

Case A, Rosen H (1993) Budget Spillovers and Fiscal Policy Interdependence. J Public Econ 52:285–307
Castro I, Santos R, Fernandes D, Xavier C, Bastos AP, Serra M (2018) O papel da terceira missão em regiões periféricas: o caso do apoio 155 da Universidade Federal do Pará ao Plano Nacional de Formação dos Professores da Educação Básica (2009–2015). In: Serra M, Rolim C, Bastos AP (eds) Universidades e Desenvolvimento Regional as Bases. Rio de Janeiro: Ideia, pp 155–188
Ceneviva R (2010) O nível de governo importa para a qualidade da política pública? A municipalização da educação fundamental no Brasil. In. Seminário discente da pós-graduação em ciência política da usp, 2. http://www.fflch.usp.br/dcp/assets/docs/SemDisc2012/07-3_Ricardo_Ceneviva.pdf. Accessed Fevereiro 2020
Da Mota ABR (2008) O processo de municipalização do ensino em Porto Felis-SP: uma análise do caminho percorrido. Sorocaba – SP. Tese de Doutorado. http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=166145. Acessed February 2020
De Dominicis M, Pinardi N, Zodiatis G, Lardner R (2013) MEDSLIK-II, a Lagrangian marine surface oil spill model for short-term forecasting—part 1: theory. Geosci Model Develop 6:1851–1869. https://doi.org/10.5194/gmd-6-1851-2013
De Andrade RJ, Soares JFO (2008) O Efeito da escola básica brasileira. Estudos em Avaliação Educacional 19(41):379–406
Diniz JA (2012) Eficiencia das transferências intergovernamentais para a educacão fundamental de municípios brasileiros. Thesis (Philosopher Degree)—University of Sao Paulo, Sao Paulo
Amaral LFE, Menezes-Filho N (2008) A relação entre gastos educacionais e desempenho escolar. In: **Anais** do Encontro Nacional de Economia, 36, 2008, Salvador. Anais. Salvador: Anpec. Available at: http://www.anpec.org.br/encontro2008
Easterly W, Levine R (1995) Africa's growth tragedy: a retrospective, 1960–1989, World Bank, Policy Research Working Paper, no. 1503
Beeson P, Montgomery E (1993) The effects of colleges and universities on local labor markets. Rev Econ Stat 75:753–761
Elhorst JP (2010) Applied spatial econometrics: raising the bar. Spatial Econ Anal 5(1):9–28
Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: theory and evidence. J Appl Econ 22(6)
Fernandes R (2013) Indicador de eficiência dos gastos na educação pública. Relatório: movimento todos pela educação, Mimeo
França MTA, de Gonçalves FO (2012) Sistemas públicos de ensino fundamental e a perpetuação da desigualdade: democracia e qualidade educacional como promotoras de justiça social. Revista Brasileira de Estudos Populacionais 29(2):303–322
Franco C, Sztajn P, Ortigão MIR (2007) Mathematics teachers, reform and equity: results from the Brazilian national assessment. J Res Math Educ Reston 38:393–419. https://doi.org/10.2307/30034880
Garcia R, Araujo V, Mascarini S (2013) The role of geographic proximity for university-industry linkages in Brazil: an emprical analysis. Austral J Regional Stud 19(3):433
Guiso L, Monte F, Sapienza P, Zingales L (2008) Culture, gender and math. Science 320(5880):1164–1165
Hanushek EA, Kimko DD (2000) Schooling, labor-force quality and the growth of nations? Am Econ Rev 90(5):1184–1208
Le HT, Nguyen HT (2018) The evolution of the gender test score gap through seventh grade: new insights from Australia using unconditional quantile regression and decomposition. J Lab Econ 7:2
Lee V (2008) Utilização e modelos hierárquicos lineares para estudar contextos sociais. In: Brooke N, Soares JF (eds) Pesquisa em eficácia escolar: origem e trajetórias. Belo Horizonte, Editora UFMG, pp 273–298
Leme M, Paredes R, Souza A (2009) A municipalização do ensino fundamental e seu impacto sobre a proficiência no Brasil. Educação básica no Brasil: construindo o país do futuro. Rio de Janeiro: Campus-Elsevier, pp 261–280
Loonis V, de Bellefon MP (coord) (2018) Handbook of Spatial Analysis: Theory and practical application with R. Insee Méthodes, No 131
Machado AF, Moro S, Martins L, Rios J (2008) Qualidade do ensino em matemática: determinantes do desempenho dos alunos em escolas públicas estaduais mineiras. Revista Economia 9(01):23–45
Magalhães A, Hewings GJD, Azzoni CR (2005) Spatial dependence and regional convergence in Brazil. Investigaciones Regionales. Madrid 6:5–20
Menezes-Filho N (2001) A Evolução da Educação no Brasil e seu Impacto no Mercado de Trabalho. São Paulo, FEA/USP

Menezes-filho N (2007) Os determinantes do desempenho escolar no Brasil. Instituto Futuro Brasil, IBMEC São Paulo e Faculdade de Economia e Administração da Universidade de São Paulo. Sumário Executivo

Moreno R, Trehan B (1997) Location and the Growth of Nations. J Econ Growth 2:399–418

Moretti E (2004) Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. J Econo 121:175–212

Murnane RJ, Willett JB, Levy F (1995) The growing importance of cognitive skill in wage determination. Rev Econ Statist 77(2):251–266

Murphy KM, Peltzman S (2004) School performance and the youth labor market. J Labor Econ 22(2):299–325

Soares JF, Alves MTG (2013) Effects of schools and municipalities in the quality of basic education. Cadernos de Pesquisa 43:492–517. https://doi.org/10.1590/S0100-15742013000200007

OECD (2019) Programme for International Student Assessment: results from PISA 2018, PISA, OECD Publishing, Paris, https://www.oecd.org/pisa/publications/PISA2018_CN_BRA.pdf

Ramos R, Suriñach J, Artís M (2010) Human capital spillovers, productivity and regional convergence in Spain. Papers Reg Sci 89(2)

Reynolds D, Teddlie, C (2008) Os processos da eficácia escolar. In: Brooke N, Soares JF (Org.). Pesquisa em eficácia escolar: Origem e trajetórias. Belo Horizonte: Ed. UFMG

Rosenthal S, Strange W (2008) The attenuation of human capital externalities. J Urban Econ 64:373–389

Silva N, Hasenbalg C (2002) Recursos familiares e transições educacionais. Cadernos de Saúde Pública 18:S67–S76

Silveira Neto RM (2001) Crescimento e spillovers: a localização importa? Evidências para os estados brasileiros. Revista Econômica do Nordeste 32:524–545

Singh M (2015) Impact of parental encouragement on the adjustment of secondary school students. Pedagogy Learn 1(1):46–52

Soares TM (2003) Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. Estudos em Avaliação Educacional 28:103–124

Soares TM (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 JF, Alves MTG, Xavier FP (2016) Effects of Brazilian schools on student learning. Assess Educ Principles Policy Pract 23(1):75–97

Vernier LDS, Bagolin IP, Jacinto PA (2015) Fatores que influenciam o Desempenho Escolar no Estado do Rio Grande do Sul: Uma Análise com Regressões Quantílicas. Análise Econômica 33(64):143–170