ACTA SCIENTIAE

ISSN: 2178-7727
DOI: 10.17648/acta.scientiae.5494

School Inequalities and Urban Welfare: Going beyond Socioeconomic Status with Data Science[1]

Renato P. dos Santos[a]
M. Şahin Bülbül[b]
Isadora L. Lemes[a]

[a] Universidade Luterana do Brasil, Programa de Pós-Graduação em Ensino de Ciências e Matemática, Canoas, RS, Brasil
[b] Kafkas Üniversitesi, Dede Korkut Eğitim Fakültesi, Kars, Türkiye

Received for publication on 31 Oct. 2019. Accepted, after revision, on 2 Nov. 2019.
Assigned editor: Claudia Lisete Oliveira Groenwald

ABSTRACT

The fast-changing Brazilian urban reality segregates people in socio-spatial terms and distributes urban public resources unfairly, threatening student's access to the structure of educational opportunities and causing school inequalities. Factors such as the existence of public lighting, open sewage and garbage accumulated around the homes, as well as electricity and water supply, sanitation, and the number of residents per bathroom, are discussed as predictors of school achievements as measured using their average IDEB (Basic Education Development Index) outcomes using Data Science methods. It was found that the resident/bathroom density and the household wall material indicators in a municipality have a higher correlation with its average school achievements than the average students' socioeconomic status, relations that are clearly illustrated through bivariate choropleth maps across all the 5,388 Brazilian municipalities with available valid data. These results are compatible with research that reveals the presence of a "neighbourhood effect," such that the distributional inequalities in infrastructure access and ultimately the notion of urban welfare reduces educational opportunities and engenders social inequalities, what is incompatible with the ideal of a sustainable society.

Keywords schooling inequalities; social inequalities; sustainability; urban inequalities; urban welfare; socio-spatial segregation

Desigualdades escolares e bem-estar urbano: indo além do status socioeconômico com a ciência de dados

RESUMO

A realidade urbana brasileira em rápida mudança segrega as pessoas em termos socioespaciais e distribui recursos públicos urbanos de maneira injusta, ameaçando o acesso do aluno à estrutura

[1] This paper is an extended version of a poster presented at the 8. Deutsch-Brasilianisches Symposium für nachhaltige Entwicklung, Porto Alegre, 2-7 Oktober 2017.

Corresponding author: Renato P. dos Santos. Email: RenatoPSantos@ulbra.edu.br

de oportunidades educacionais e causando desigualdades na escola. Fatores como a existência de iluminação pública, esgoto a céu aberto e lixo acumulado nas residências, além de fornecimento de eletricidade e água, saneamento e o número de moradores por banheiro, são discutidos como preditores de realizações escolares, medidos usando seu IDEB médio (Índice de Desenvolvimento da Educação Básica) usando métodos de Ciência de Dados. Verificou-se que os indicadores de densidade de residentes/banheiros e de material da parede do domicílio em um município têm uma correlação mais alta com o desempenho escolar médio do que o status socioeconômico médio dos alunos, relações que são claramente ilustradas por meio de mapas coropléticos bivariados sobre todos os 5.388 municípios brasileiros com dados válidos disponíveis. Esses resultados são compatíveis com pesquisas que revelam a presença de um "efeito de vizinhança", de modo que as desigualdades distributivas no acesso à infraestrutura e, finalmente, a noção de bem-estar urbano reduzem as oportunidades educacionais e geram desigualdades sociais, o que é incompatível com o ideal de uma sociedade sustentável.

Palavras-chave: desigualdades na escolaridade; desigualdades sociais; sustentabilidade; desigualdades urbanas; bem-estar urbano; segregação socioespacial

INTRODUCTION

According to Soares (2004), the factors that determine students' cognitive performance can be grouped into three broad categories: those associated with school structure, those associated with the family, and those related to the student itself.

Empirical research conducted from the 1950s to the 1990s in the United States, England, and France showed that extra-school factors explain better the observed inequalities in student performance than intra-school factors. In particular, they showed that both access to education and school outcomes are mostly and strongly directly associated with the socioeconomic and cultural characteristics of the students (Soares, 2004).

The economic growth of a society does not automatically translate into an improvement in the quality of life, environmental preservation, or universal quality education but often in a strengthening of inequality (PNUD, IPEA, & FJP, 2014).

The process of countries' economic development is characterised by the movement of people from rural to urban areas, increasing its rate of urbanisation (Castells-Quintana, 2018).

Brazilian urbanisation presented a mismatch about the economic growth when it began to be driven by the industrialisation process, differently from the experiences of the countries during the first industrialisation, where there was more correspondence (Ribeiro, 2016). The uncontrolled growth of population and cities in Brazil segregates people in socio-spatial terms, distributes urban public resources unfairly and consequently threatens student's access to the structure of educational opportunities (de Queiroz Ribeiro, Koslinski, Zuccarelli, & Christovão, 2016) (Ribeiro, Koslinski, Zuccarelli, & Christovão, 2016).

Aware of this reality, the Human Development Index (HDI), was idealized as a measure of the degree of human development in a country, as an alternative to the Gross Domestic Product (GDP), by the Pakistani economist Mahbub ul Haq, with the help of economist Amartya Sen (United Nations Development Programme (UNDP), 1990).

Along similar lines, the Observatory of the Metropolises (*Observatório das Metrópoles*) developed the Urban Welfare Index (IBEU), having five dimensions related to urban mobility, urban environmental conditions, urban housing conditions, attendance of collective public services, and the existence of urban infrastructure. Each one of these dimensions comprises one to seven indicators, built from the demographic 2010 census from the Brazilian Institute of Geography and Statistics (IBGE) (de Queiroz Ribeiro & Ribeiro, 2013).

However, for decades, the economic literature on infrastructure has mostly focused on aggregate data at the regional or more often, the national level. Traditionally, infrastructure has either been measured through supply-side physical indicators, such as electricity generation capacity or kilometres of roads, or demand-side ones such as aggregate electricity or water connection rates. Given the fact that such data ignore both the spatial nature of and the distributional disparities in infrastructure access, the policy relevance of the conclusions have been limited (Fay & Straub, 2017, p. 3) (see Straub (2011) for a critical review).

Wei et al. (2018) confirm not only the significance of neighbourhood context in students' academic performance but also the importance of integrating GIS spatial analysis tools into studying education inequality.

Therefore, in this research, we expanded the study by de Queiroz Ribeiro et al. (2016)(2016) about Brazilian metropolises to all the 5,564 Brazilian municipalities to investigate to a higher level of detail the relation between residential segregation and distributional disparities in infrastructure access and school inequalities.

BACKGROUND

The relation between school outcomes and the conditions outside but nearby it have been studied for some time. The effects of segregation on socio-spatial mobility reduction in Sweden, the Netherlands, the UK (England and Wales), and Estonia was studied by Nieuwenhuis et al. (2019). The neighbourhood effects and parental characteristics on school achievement in the Netherlands were studied by Nieuwenhuis et al. (2013). Friedman (2018) studied the relation between fast urbanisation growth and schooling in China. Umar (2017) examined how variations in urbanisation and household size impact on educational inequality in Nigeria. Cameron (2017) analysed the relations between urban inequality, social exclusion and schooling in Bangladesh. Tan, Ho, & Pang (2015) compared education inequalities in urban and rural areas in Malaysia. Lewis-Mccoy (2014) discusses segregation, social mobility, and educational opportunity among US

African Americans. De Bruin and Liu (2019) analysed urbanisation, gender and schooling inequality in China.

According to Mack (2017), cities have always been unique places and opportunities for not only formal but also non-formal and informal education, thanks to their central spatial role with regard to their surroundings. However, Mack (2017) also points out that there are also conflicts of interest and social disparity on the urban policy agenda, as well as broken promises and unresolved conflicts.

Parents tend to choose their residence based on neighbouring school quality, often leaving or avoiding areas with higher levels of deprivation, therefore often reproducing residential segregation in school segregation, which in turn can lead to labour market segregation and socio-spatial mobility reduction (Nieuwenhuis et al., 2019).

On the other hand, Vetter et al. (1981) argues that this urban segregation and unequal distribution of urban collective resources is powered from a "circular causation," in which the highest income social groups, benefited by the actions of the State in a given period that increase their real income by the valorisation of the soil price, which impacts on the residential segregation and prevents the entry of lower-income social groups in those spaces, which further increases their power to claim for future distributions of the net benefits of state actions.

It has been argued that the existing social, economic and working conditions, apparently external to the school, are nonetheless present in its day-to-day experience (Carter, 2016; Marqués Perales, 2016). In other words, as Boudon argued before (1979), schools alone cannot eradicate inequality if the needs of the environmental community are not addressed (Carter, 2016).

However, we are not subscribing here the strong pedagogical pessimism that followed the publication of Coleman Report (1966), usually synthesised in the provocative phrase "schools make no difference" [about inequality], popularly associated with Coleman's work. As a matter of fact, various authors argue that Coleman Report has generally not been read in its 737 pages entirety, and has been summarised in a biased and simplistic way (Martín-Lagos López, 2018).

Quite on the contrary, Coleman believed that school inequality could make things worse, per Downey and Condron (2016) understanding that schools can reproduce inequality, compensate for it or exacerbate it. Coleman Report indicates that one student's achievement appears to be "strongly related to the educational backgrounds and aspirations of the other students in the school," in the sense that the higher the social class of other students the higher any given student's achievement, with the opposite also being true, independently of the student's own social background. For example, according to Coleman Report, individuals who might never consider dropping out if they were in a different high school might decide to drop out if they attended a school with many low-income colleagues, who drop out at substantially higher rates than middle-class students do. Furthermore, Coleman's survey found that students were 14 times as likely to say it was harder to accept the disapproval of peers than of teachers. As Coleman flatly declared, "A child's learning is a function more of the characteristics of his classmates than those of the teacher."

There are both direct and indirect effects of the economic conditions on students' cognitive performance. Indirect effects are especially significant as economic conditions also act to create unique conditions for the consumption of cultural goods and provide parents with the time necessary to follow the children's school life (Soares, 2004).

Therefore, as (Harvey, 1971)the students' socioeconomic status does not account for the complexity of these indirect effects of the economic conditions, de Queiroz Ribeiro & Ribeiro (2013, p. 10) resorted to David Harvey's concept of 'real income', introduced in his seminal book *Social Justice and the City* (1971). This concept surpasses the "simplest, and perhaps misleading," definition of income as "the amount received in spendable form in a given year" (1971, p. 53) and include non-monetary "fringe benefits" (1971, p. 54) that shall be provided by the city and used collectively by the people, in terms of material conditions of life, in the form of housing, transportation, and quality education.

Still, according to Harvey, the city can be thought of as a gigantic resource system that contains human-made resources of great economic, social, psychological, and symbolic significance. It is also a geographically localised system in the sense that most of those resources we make use of "are nor ubiquitous and their availability, therefore, depends upon accessibility and proximity" (1971, pp. 68–69). Furthermore, resources are also technological and cultural appraisals – their quantity is dependent upon the individual preferences existing in the population and the cognitive skills which people possess to help them exploit the resource system (1971, p. 69).

According to de Queiroz Ribeiro & Ribeiro (2013, p. 11), the advantage of using the concept of real income to define urban welfare is due to the fact that the collective resources existing in contemporary society, which can contribute to the improvement of living conditions, are distributed unevenly in the metropolis. That is, this concept enables us to evaluate how urban conditions favour social inequalities, as public resources are unevenly distributed among social groups across the city.

Consequently, these concepts of real income and urban welfare allow one to evaluate better, how the uneven distribution of urban conditions, e.g. access to quality education, among social groups in the city favours social inequalities and may be a better (Ribeiro & Ribeiro, 2013, pp. 9–11)predictor of their school achievements.

METHODOLOGY

The so-called 'data mining' was until recently considered one of the five stages of the *Knowledge Discovery in Databases* (KDD) process proposed by Fayyad, Piatetsky-Shapiro, and Smyth (1996), which aimed at identifying new, valid, potentially useful and understandable patterns that were embedded in the data. These stages were the selection of adequate bases, the cleansing of the inconsistencies that can afflict this data, the transformation of the data into more adequate formats, the choice of techniques and algorithms for mining, and the evaluation and interpretation of the extracted patterns in the form of the new knowledge. Today, the first three steps are grouped in a phase called *exploratory data analysis* (Peng, 2016).

As said before, the Observatory of the Metropolises (*Observatório das Metrópoles*) developed the Urban Welfare Index (IBEU) (de Queiroz Ribeiro & Ribeiro, 2013), having five dimensions (D1 – D5), each composed of one to seven indicators (Table 1).

Table 1
Dimensions and indicators composing the Urban Welfare Index (IBEU)

D1 – Urban Mobility	
	Work-to-home travel time
D2 – Urban Environmental Conditions	
	Forestation around households
	Open sewage in the vicinity of households
	Accumulated garbage around households
D3 – Urban Housing Conditions	
	Subnormal clump
	Resident/dormitory density
	Resident/bathroom density
	Adequacy of household wall material
	Adequacy of households
D4 – Attendance of Urban Collective Services	
	Water supply
	Sewage treatment
	Electricity supply
	Waste collection
D5 – Urban Infrastructure (Existence of)	
	Public lighting
	Paving
	Sidewalk
	Curb / Guide
	Road gully
	Wheelchair ramp
	Street address identification

Note: de Queiroz Ribeiro & Ribeiro (2013)

Another option would be using the MHDI, a version of the Human Development Index (HDI) (United Nations Development Programme (UNDP), 1990) proposed by Permanyer (2013) that allows exploring the detailed distribution of human development to the municipal level. However, Pinheiro et al. (2016) have shown that the MHDI and

IBEU indicators have equivalent powers in determining the development of a locality, while not dismembered in their individual components.

Although each one of the proposed dimensions deals with a particular type of urban well-being, IBEU reliability in the sense that the measures of its dimensions for the 15 Brazilian metropolises are related to each other was verified through the *Cronbach Alpha coefficient* test by de Queiroz Ribeiro, and Ribeiro (2013, p. 29). The obtained result of 0.750 proved IBEU robust enough to express the conditions of urban well-being at metropolises level, although being quite below the desirable value one also indicates that those dimensions are not sufficient to capture all its aspects (de Queiroz Ribeiro & Ribeiro, 2013, p. 29). Furthermore, the dimension work-to-home travel time (D1), which is used as a proxy to urban mobility, has a noticeable minor correlation with the others (de Queiroz Ribeiro & Ribeiro, 2013, pp. 29–30). This result is compatible with the comparative analysis between IDH-M and IBEU done by Pinheiro et al. (2016), in which it was observed no statistically significant correlation between the HDI-M and the dimension D1.

The intention here was to expand the study by de Queiroz Ribeiro et al. to investigate down to the Brazilian municipalities' level the relationship between residential segregation and distributional disparities in the access of infrastructure and school inequalities, using Data Science methods.

For indicators of school achievements, we used the ENEM (National High School Exam) 2015 outcomes and IDEB (Basic Education Development Index) 2011 values averaged across each municipality. Again, the 'Education' component of HDI could have been considered; however, it must be noticed that it actually is "a measure of deprivation that a country suffers in" the primary variable 'literacy' (United Nations Development Programme (UNDP), 1990, p. 109), calculated from the 'adult literacy rate'[2] in that country.

From 2014 on, the ENEM outcomes databases also started providing information about the socioeconomic status of the schools to contextualise the indicators of school achievements (Brasil. INEP, 2015). This measure of socioeconomic level is calculated from data on the possession of household goods, income and contracting services by the students' family and their parents' educational level. It is expressed in seven levels: Very Low, Low, Medium Low, Medium, Medium High, and Very High (Brasil. INEP, 2015). This indicator was calculated for each school, averaged across all schools in each municipality, and rescaled to the range 0–1 to facilitate comparison with the other indices.

Firstly, we calculated those IBEU dimensions and indexes from the Brazilian 2010 Population Census (IBGE, 2012) data (2016)for all municipalities, as indicators of residential segregation and distributional disparities in the access to infrastructure.

[2] Defined as the percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life

Secondly, we analysed those IBEU dimensions and indexes in detail, looking for spatial inequalities across the individual municipalities employing measures of inequality, such as the *Gini coefficient*, and measures of spatial autocorrelation, such as *Moran's I* statistic, first proposed by Moran (1948). As usual, near-zero values for Gini coefficient and Moran's I statistics indicate perfect equality and no spatial autocorrelation (random spatial distribution), respectively, while positive values indicate positive spatial autocorrelation, with spatial clusters of similarly low or high values between neighbour municipalities, and negative values indicate negative spatial autocorrelation, in which low values tend to have neighbours with high values and vice versa.

However, it must be noticed that classical Moran's I statistic and Gini coefficient are both whole-map, locationally invariant measures, in the sense that they can tell whether something is happening, but not where it is happening within the region of interest (Rey & Smith, 2013). Therefore, we also used a spatial decomposition of the Gini coefficient introduced by Rey and Smith (2013) that supports the detection of spatial autocorrelation and segregation, as provided by the *lctools* package (Kalogirou, 2019). Furthermore, the *lctools* package also provides a Monte Carlo simulation, in which the data are spatially reallocated in a random way to infer the share of overall inequality that is associated with non-neighbour pairs of locations and, therefore, to assess the significance of the evaluated Spatial Gini coefficient (Rey & Smith, 2013).

Thirdly, we investigated eventual correlations between IDEB and ENEM and each one of those IBEU dimensions and indicators, firstly as whole-map statistics, while controlling for the student's socioeconomic status, by means of *Partial Pearson's Correlation Coefficients* (PPCC), as made available by Kim through his *ppcor* package (2015a, 2015b).

Fourthly, we proceeded to analyse those correlations between IDEB and each one of those IBEU dimensions and indicators across each municipality. We started by drawing scatterplots showing IDEB and IBEU indexes and Socioeconomic Status index values for each municipality and confidence ellipses, based on multivariate *Student-t* distributions, with a standard confidence level of 0.95, making use of the *ggplot2* package (Wickham & Chang, 2019, Chapter stat_ellipse) for the R statistical data analysis language (R Core Team, 2019). When the confidence ellipse is 'tilted,' the explanatory variables are correlated; in contrast, when its axes are parallel to the axes of the parameter space, the explanatory variables are uncorrelated (Fox, 2016, p. 221).

In the following, as such "global" statistics are likely to hide significant spatial variation of the relationship between two variables (Kalogirou, 2015), we moved to using *Local Pearson's Correlation Coefficients* (LPCC) based on a fixed number of nearest neighbours, as proposed by Kalogirou (2012, 2013) and provided by the *lctools* package (Kalogirou, 2019). As argued by Kalogirou (2012), these LPCC would allow the identification of pairs of variables that not significantly correlated globally but higher correlated locally. The *lctools* package also provides a Monte Carlo simulation proposed by Hope (1968) and adapted by Fotheringham, Brunsdon, and Charlton (2002) to assess whether the spatial variation of the local correlation coefficients is statistically significant.

Finally, we analysed those spatial inequalities visually across all the individual municipalities with *choropleth maps*. Choropleth maps are colour-coded geographic maps in which thematic values, such as demographic density or per capita income, are proportionally coded using some smooth colouring function to RGB domain and then applied as patterns or shadings to particular geographic areas of occurrence, typically administrative regions. Any singular (univariate) values related to the event can then be so projected, and the viewer can quickly get an impression of its distribution across the regions within the map. A more complex, bivariate choropleth map, as provided by the *colorplaner* package (Murphy, 2016), uses a smooth colouring function that accounts for two-dimensional variables and is suited to identifying correlations between those variables from the single displayed colour. Positive correlations between the variables are indicated by colours ranging from green (both variables have low values) to violet (both high) while colours such as red and blue indicate negative correlations (one is high, and the other is low or vice-versa).

Except for the choropleth maps, all the other graphs were built using the *ggplot2* package developed by Wickham (Wickham, 2016) and further enhanced by (Wickham & Chang, 2019).

RESULTS

We calculated those IBEU dimensions and indexes from Brazilian 2010 Population Census (IBGE, 2012) data (2016)for all the 5,564 municipalities, as indicators of residential segregation and distributional disparities in the access of infrastructure, resulting in valid data for 5,388 municipalities only.

As ENEM is a voluntary membership assessment available to all high school graduates or people already trained at this level (Andrade & Soida, 2015), from the 2,085,245 students enrolled in the 3rd year of Regular High School, as declared in the 2015 Basic Education Census, 1,385,394 enrolled in ENEM 2015 and only 1,212,908 students took all the ENEM tests, receiving a grade higher than zero on the outcomes and, therefore were not eliminated from the assessment (Brasil. INEP, 2016). Furthermore, according to the ENEM criteria, schools with less than ten students and/or less than 2% (of enrolled students) participating in the exam are excluded from the listings for that exam (Andrade & Soida, 2015). Consequently, there is valid ENEM data available for 3,784 municipalities only. Consequently, there is data on the socioeconomic status of the schools available for those same 3,784 municipalities only.

Analysing those indicators, looking for spatial inequalities across all the individual municipalities quantitatively by means of measures of segregation, such as the Spatial Gini coefficient, results from Monte Carlo simulations for the significance of the Spatial Gini coefficient, and measures of degrees of clustering, such as Moran’s I statistics, the corresponding Expected Moran's I values under the null hypothesis of no spatial autocorrelation, *z-scores*, and (two-tailed) *p*-values calculated for the resampling and randomization null hypotheses tests, as provided by the *lctools* package (Kalogirou, 2019) for the R language.

As seen from Table 2, even though the values of the overall Gini are not close to 1, they are still significant, according to its spatial decomposition, which shows that most of them $(ns.frac \sim 1)$ come from spatial autocorrelation between non-neighbour municipalities and clustering, allowing us to discard virtually any spill-over effects. Furthermore, all *z-score* values, both for resampling and randomisation null hypotheses tests, are much higher than 1.96, which, under the assumption of a normal distribution, correspond to the extremely small p-values displayed. These results indicate a high grade of inequality on the factors represented by these indicators across Brazilian municipalities, as expected.

Table 2
Measures of segregation and clustering for the indicators across the individual municipalities.

Indic.	Gini	gw.frac	ns.frac	*p*	Moran.I	EI	z.res	z.rand	p.rsamp	p.rand
D1	0.03	2.1E-03	0.998	0.05	0.55	-2.7E-04	86.45	86.68	0.0E+00	0.0E+00
D2	0.08	1.9E-03	0.998	0.05	0.55	-2.7E-04	86.09	86.10	0.0E+00	0.0E+00
D3	0.05	1.2E-03	0.999	0.05	0.82	-2.7E-04	128.21	128.25	0.0E+00	0.0E+00
D4	0.15	1.5E-03	0.998	0.05	0.71	-2.7E-04	110.71	110.70	0.0E+00	0.0E+00
D5	0.14	1.8E-03	0.998	0.05	0.58	-2.7E-04	91.11	91.11	0.0E+00	0.0E+00
SE	0.17	1.4E-03	0.999	0.05	0.72	-2.7E-04	112.94	112.93	0.0E+00	0.0E+00
IDEB	0.10	1.7E-03	0.998	0.05	0.66	-2.7E-04	103.37	103.36	0.0E+00	0.0E+00
ENEM	0.03	2.3E-03	0.998	0.05	0.39	-2.7E-04	60.92	60.93	0.0E+00	0.0E+00
IBEU	0.07	9.3E-04	0.999	0.05	0.77	-1.9E-04	145.37	145.37	0.0E+00	0.0E+00

Note: *gw.frac* and *ns.frac* are the shares of overall Gini that are associated with neighbours and non-neighbours, respectively, and *p* is the value of the corresponding test statistic. *EI* is the expected value of *Moran's I* under the null hypothesis of no spatial autocorrelation. *z.res*, *z.rand*, *p.rsamp*, and p.rand are the z-scores and test statistics for resampling and randomisation null hypotheses tests.

The correlation analysis summarised in Table 3 shows some correlation between students' Socioeconomic Status and most of the indicators that compose IBEU, as well as with the ENEM and IDEB indicators. For this reason, it seems reasonable to control this variable while doing correlation analysis of IDEB with the indicators and dimensions composing IBEU.

Table 3
PCC Correlations between Socioeconomic Status and other indicators

Indicator	Dim.	*r*	*p*
Resident/bathroom density	D3	0.69	0.00E+00
ENEM	–	0.68	0.00E+00
Waste collection	D4	0.65	0.00E+00
Road gully	D5	0.54	0.00E+00

Indicator	Dim.	*r*	*p*
IDEB	–	0.48	0.00E+00
Resident/dormitory density	D3	0.46	0.00E+00
Adequacy of household wall material	D3	0.42	0.00E+00
Sewage treatment	D4	0.41	0.00E+00
Open sewage	D2	0.40	0.00E+00
Electricity supply	D4	0.33	0.00E+00
Address identification	D5	0.27	0.00E+00
Public lighting	D5	0.23	0.00E+00
Paving	D5	0.18	0.00E+00
Curb / Guide	D5	0.18	0.00E+00
Water supply	D4	0.15	0.00E+00
Sidewalk	D5	0.12	5.43E-13
Accumulated garbage	D2	0.08	9.98E-07
Adequacy of Home-to-Work Displacement	D1	-0.07	1.18E-05
Adequacy of households	D3	NS	0.33
Wheelchair ramp	D5	NS	0.18
Subnormal clump	D2	NS	0.11
Forestation around households	D2	NS	0.11

Note: 'NS' indicates a non-statistically significant correlation ($p > 0.5$).

Furthermore, Table 3 also shows that students' socioeconomic status correlates quite better with ENEM than with IDEB. For this reason, and because of the above-mentioned lack of data on ENEM outcomes for many municipalities, we will focus solely on IDEB as the indicator of school achievements.

Partial correlation analysis while controlling for the Socioeconomic Status variable indicates weak aggregate correlations between IDEB and all the indicators (Table 4). Notably, the correlation *IDEB vs ENEM* changes from $r = 0.35$ with $p = 1.8 \times 10^{-104}$ to not statistically significant ($p = 0.06$) when we control for the Socioeconomic Status variable; this comes from the above-mentioned influence of the Socioeconomic Status on ENEM.

Table 4
PCC, PPCC, and SPPCC values between IDEB and the indicators composing IBEU while controlling for the Socioeconomic Status variable

Indicator	Dim.	*r*	*p*	*r* (partial)	*p*	*r* (semi-partial)	*p*
Open sewage	D2	0.42	8.0E-159	0.29	3.6E-70	0.25	4.8E-54
Resident/bathroom density	D3	0.51	2.1E-245	0.28	5.2E-69	0.25	3.7E-53
Resident/dormitory density	D3	0.42	4.3E-161	0.26	3.6E-59	0.23	1.1E-45
Paving	D5	0.30	8.6E-79	0.25	4.8E-52	0.22	2.8E-40
Curb / Guide	D5	0.28	1.2E-66	0.22	2.2E-42	0.20	6.6E-33
Home-to-work travel time	D1	0.14	7.4E-18	0.20	6.4E-35	0.18	3.3E-27

Indicator	Dim.	*r*	*p*	*r* (partial)	*p*	*r* (semi-partial)	*p*
Road gully	D5	0.40	2.0E-138	0.19	1.9E-30	0.16	8.9E-24
Street address identification	D5	0.29	2.3E-73	0.19	1.1E-31	0.17	9.7E-25
Household wall material	D3	0.34	2.9E-103	0.18	1.1E-29	0.16	3.5E-23
Accumulated garbage	D2	0.20	3.0E-33	0.18	6.7E-28	0.16	8.1E-22
Electricity supply	D4	0.28	3.6E-69	0.16	1.0E-21	0.14	4.6E-17
Sewage treatment	D4	0.30	2.6E-77	0.13	1.4E-14	0.11	1.4E-11
Water supply	D4	0.18	2.7E-29	0.12	3.2E-14	0.11	2.7E-11
Sidewalk	D5	0.13	2.9E-15	0.08	4.0E-07	0.07	8.6E-06
Forestation	D2	0.07	8.8E-06	0.07	3.6E-05	0.06	2.8E-04
Subnormal clump	D3	0.05	2.6E-03	0.07	7.7E-05	0.06	5.2E-04
Public lighting	D5	0.14	1.7E-18	0.04	1.5E-02	0.04	3.2E-02
Adequacy of households	D3	NS	0.26	0.03	3.6E-02	NS	6.6E-02
ENEM	–	0.35	1.8E-104	NS	0.06	NS	0.10
Waste collection	D4	0.32	8.2E-91	NS	0.35	NS	0.41
Wheelchair ramp	D5	NS	0.36	NS	0.62	NS	0.67

Note: 'NS' indicates a non-statistically significant correlation ($p > 0.5$).

Not unexpectedly, partial correlation analysis between IDEB and all the aggregated dimensions of IBEU while controlling for the Socioeconomic Status variable also results in weak correlations (Table 5).

Even so, *Open sewage (D2)* is the individual indicator with the highest correlation with IDEB, even while controlling for the Socioeconomic Status variable.

This result is coherent with Fay and Straub research, which indicates that the water cluster explains between 40 and 67% of the variability of individual percentile-level access to infrastructure services and consumption of related assets across Latin America (Fay & Straub, 2017, p. 9).

An explanation for this correlation may come from a systematic review of the literature by Jasper, Le, and Bartram (2012) on the effects of water and sanitation in schools. This review reports an increase in absenteeism from schools in developing countries due to inadequate sanitation facilities and consequently diarrheal and gastrointestinal diseases.

Table 5
PCC, PPCC, and SPPCC between IDEB and the dimensions composing IBEU while controlling for the Socioeconomic Status variable

Dimension	*r*	*P*	*r* (partial)	*p*	*r* (semi-partial)	*p*
Urban Housing Conditions (D3)	0.50	5E-231	0.30	3.2E-76	0.26	1E-58
Urban Infrastructure (D5)	0.38	8E-128	0.25	8.7E-54	0.22	1E-41
Urban Environmental Conditions (D2)	0.32	1E-90	0.23	2.4E-47	0.21	1E-36
Urban Mobility (D1)	0.14	7E-18	0.20	6.4E-35	0.18	3E-27
Urban Collective Services (D4)	0.36	7E-111	0.14	4.1E-18	0.12	3E-14

These low values for the correlation coefficients both for the indicators and the dimensions composing the IBEU demonstrate the above-mentioned limited policy relevance of aggregate data at the national level in a country with as many contrasts as Brazil (Fay & Straub, 2017, p. 3).

Moving on to a detailed analysis across the individual municipalities, Figure 1 displays scatterplots showing IDEB and IBEU indexes and *Socioeconomic Status* index (SE) values for each municipality. As the major axis of the confidence ellipse in the scatterplot for *IDEB vs D1* is almost parallel to the y-axis of the graph, a very weak correlation is to be expected between these variables; conversely, as the confidence ellipse for *IDEB vs D5* is the most 'tilted' one, this pair of variables is expected to be the best-correlated one (Fox, 2016, p. 221).

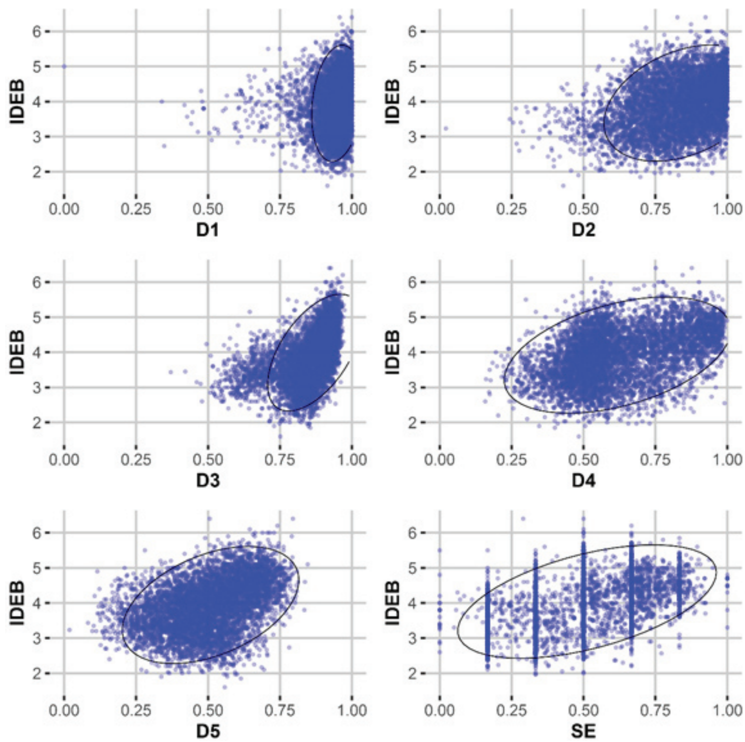

Figure 1. Scatterplots showing IDEB and IBEU indexes and Socioeconomic Status index (SE) values for each municipality.

When we move to the LPCC statistics across each municipality, we observe from Table 6 that indeed while the relation *IDEB vs D1* was the one with the highest share of statistically significant ($p \leq 0.5$) correlations, it was also the one with the shortest range of correlation coefficients.

Table 6

Shares and ranges of statistically significant ($p \leq 0.5$) correlations, as measured by LPCC values between IDEB and the dimensions composing IBEU

Dimensions	n	% of $p \leq 0.5$	*r* ($p \leq 0.5$)
Urban Mobility (D1)	5388	91.7%	0.10-0.27
Urban Environmental Conditions (D2)	5388	65.0%	0.10-0.46
Urban Housing Conditions (D3)	5388	74.0%	0.10-0.65
Urban Collective Services (D4)	5388	71.3%	0.10-0.65
Urban Infrastructure (D5)	5388	69.5%	0.10-0.55
Socioeconomic Status (SE)	3686	64.3%	-0.15-0.51

To get a better grasp of these distributions of local correlation coefficients across the municipalities, the histograms on Figure 2 and the averages, standard deviations, kurtoses, and asymmetries of the distributions in Table 7 provide further insights.

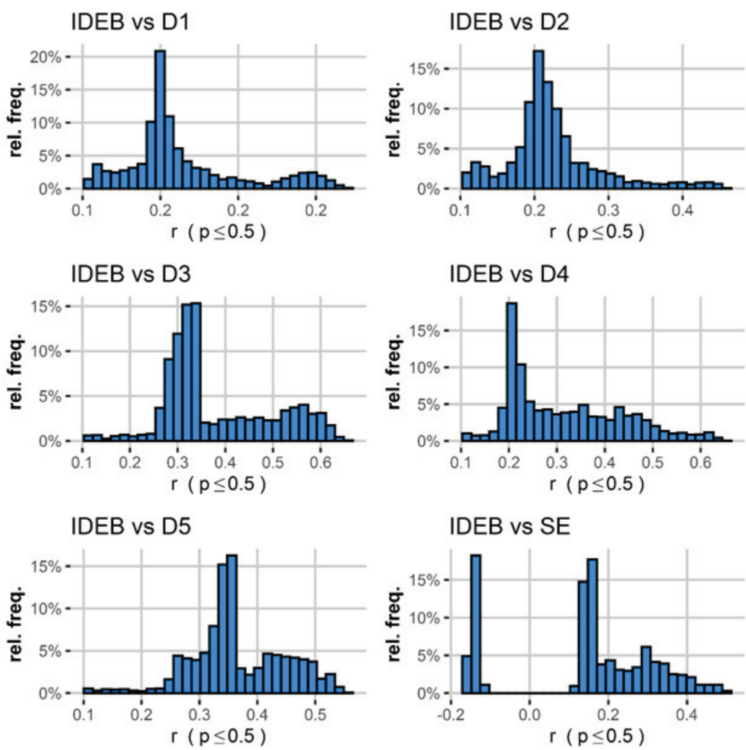

Figure 2. Histograms of Local Pearson's Correlation Coefficients (LPCC) across the municipalities.

The bimodal distribution of local correlation coefficients for the relation of IDEB with the *Socio-Economic Status* index stands out from Figure 2, including both positive and negative values, resulting in the lowest average value in Table 7; this was already observable from Table 6 and makes the value of 0.13 for its average quite meaningless.

Table 7
Averages, standard deviations, kurtoses[3], and asymmetries of the LPCC values between IDEB and the dimensions composing IBEU across the municipalities

Dimensions	avg	stdev	kurt	skew
Urban Housing Conditions (D3)	0.38	0.12	-0.49	0.50
Urban Infrastructure (D5)	0.36	0.08	0.51	-0.23
Urban Collective Services (D4)	0.32	0.12	-0.49	0.65
Urban Environmental Conditions (D2)	0.22	0.07	1.95	1.00
Urban Mobility (D1)	0.16	0.04	0.40	0.91
Socioeconomic status (SE)	0.13	0.18	-0.88	-0.33

From the scatterplots on Figure 2 and the results on Table 7, we also observe that indeed the relation *IDEB vs D1* is the one with the lowest range of correlation coefficients, with a distribution centred around the relatively small value 0.2, the smallest values for average (0.16) (disregarding the corresponding value for *IDEB vs SE*) and standard deviation (0.04). Furthermore, while the distributions of the coefficients for the correlations with the remaining dimensions of IBEU extend up to relatively larger values such as $r = 0.6$, the relation *IDEB vs D5* is the one that is more centred around a higher value, with average value of 0.36 and standard deviation of 0.08, what suggests a more definite correlation between those variables across the municipalities, as expected from the scatterplot on Figure 2. From the results on Table 7, one also observes that all those distributions are far from normal ones, being the relation *IDEB vs D5* the less skewed one (– 0.23) while still quite heavy-tailed, with an excess kurtosis value of 0.51.

Similarly, we calculated the shares and ranges of statistically significant ($p \leq 0.5$) correlations, as measured by LPCC values (Table 8), as well as their averages, standard deviations, kurtoses, and asymmetries (Table 9), for the dimensions composing IBEU across each municipality.

Table 8
Shares and ranges of statistically significant ($p \leq 0.5$) correlations, as measured by LPCC values between IDEB and dimensions composing IBEU

Indicators	Dim.	n	% of $p \leq 0.5$	r ($p \leq 0.5$)
Resident/bathroom density	D3	5388	75.0%	0.10-0.65
Open sewage	D2	5388	92.1%	0.10-0.49
Resident/dormitory density	D3	5388	43.5%	0.10-0.57
Road gully	D5	5388	62.6%	-0.21-0.39

[3] As it is common practice, to simplify the comparison with the normal distribution, we use here the so-called "excess kurtosis", calculated by subtracting 3 from the Pearson kurtosis.

Indicators	Dim.	n	% of $p \leq 0.5$	r ($p \leq 0.5$)
ENEM	–	3686	61.0%	0.12-0.39
Household wall material	D3	5388	70.6%	0.10-0.58
Waste collection	D4	5388	64.6%	-0.13-0.58
Paving	D5	5388	69.7%	0.10-0.47
Sewage treatment	D4	5388	70.2%	-0.12-0.59
Street address identification	D5	5388	91.8%	0.10-0.42
Curb / Guide	D5	5388	68.7%	0.11-0.51
Electricity supply	D4	5388	71.8%	0.10-0.46
Accumulated garbage	D2	5388	61.9%	0.10-0.22
Water supply	D4	5388	42.8%	0.10-0.32
Home-to-work travel time	D1	5388	91.7%	0.10-0.27
Public lighting	D5	5388	52.7%	0.10-0.38
Sidewalk	D5	5388	71.3%	-0.18-0.53
Forestation	D2	5388	43.8%	0.10-0.37
Subnormal clump	D3	5388	0.0%	–
Adequacy of households	D3	5388	9.2%	-0.13-0.14
Wheelchair ramp	D5	5388	0.0%	–

Table 9
Averages, standard deviations, kurtoses[4], and asymmetries of the LPCC values between IDEB and the indicators composing IBEU across the municipalities

Indicator	Dim.	avg	stdev	kurt	skew
Resident/bathroom density	D3	0.36	0.12	-0.67	0.54
Household wall material	D3	0.34	0.08	4.17	0.08
Paving	D5	0.34	0.07	0.61	-0.40
Curb / Guide	D5	0.33	0.08	-0.16	0.13
Waste collection	D4	0.30	0.12	3.24	0.14
Electricity supply	D4	0.26	0.09	2.19	0.71
Sidewalk	D5	0.26	0.16	4.65	-1.44
Resident/dormitory density	D3	0.25	0.13	-0.88	0.53
Open sewage	D2	0.22	0.08	0.22	0.72
Street address identification	D5	0.21	0.08	2.32	0.47
ENEM	–	0.21	0.07	5.80	0.20
Waste collection	D4	0.19	0.15	3.44	-0.17
Forestation	D2	0.17	0.06	5.03	1.28
Water supply	D4	0.17	0.04	4.01	0.91
Public lighting	D5	0.17	0.06	4.05	1.16
Home-to-work travel time	D1	0.16	0.04	0.40	0.91
Accumulated garbage	D2	0.14	0.02	3.33	0.22
Road gully	D5	0.07	0.18	1.73	-0.27
Adequacy of households	D3	0.06	0.08	3.40	-1.52
Wheelchair ramp	D5	-0.09	0.00	1.84	-0.01
Subnormal clump	D3	–	–	–	–

[4] As it is common practice, to simplify the comparison with the normal distribution, we use here the so-called "excess kurtosis", calculated by subtracting 3 from the Pearson kurtosis.

From the results of Table 9, one observes that the distributions of values for most indicators are far from normal. Notably, *ENEM*, *Forestation*, *Sidewalk*, and *Household wall material* were the most heavy-tailed ones. On the other hand, *Adequacy of households*, *Sidewalk*, and *Forestation* are the most skewed ones.

It is noticeable how various indicators with the highest average *PPCC* and *SPPCC* values in Table 4, such as *Open sewage* and *Resident/dormitory density*, occupy quite lower positions in Table 9 due to their smaller average *LPCC* values, while others, such as *Household wall material*, raised positions expressively, reinforcing the discussion above on the relevance of localised data over limited analysis upon aggregate ones (Fay & Straub, 2017, p. 3). Coherently, however, the indicators at the lowest positions in Table 4 also occupy the bottom positions in Table 9. Nevertheless, it is worth mentioning *Resident/bathroom density* and *Paving* as the only somewhat top indicators in Table 4 that kept top positions in Table 9, in the same order.

The scatterplots on Figure 3 for a few top indicators from Table 9 across each municipality, as well as the results on Table 9 itself, show that the values of most of these indicators do not have a normal distribution across the municipalities but rather suffer from an accumulation near the maximum value. This shows that they have limited utility for correlations with normalised indicators such as IDEB. Furthermore, this issue affects the dimensions of IBEU, as seen from the scatterplots in Figure 2 and the results in Table 7.

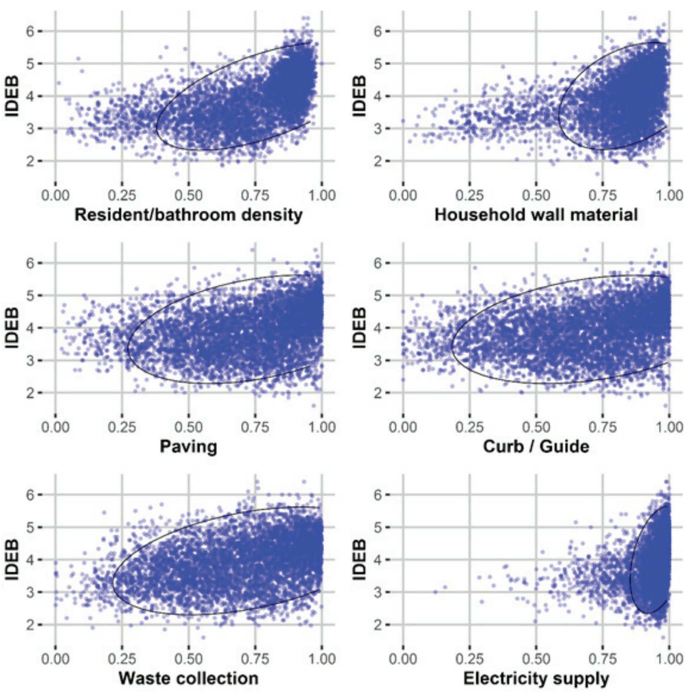

Figure 3. Scatterplots showing IDEB and the top six indicators from Table 9 values for each municipality.

To appreciate better those distributions for the dimensions that compose IBEU, we proceed to build choropleth maps to analyse those spatial inequalities visually across all the individual municipalities. However, instead of applying those local correlation coefficients to the particular municipalities they correspond, we built bivariate choropleth maps for the variables themselves, leaving the correlations between them to be identified visually directly from the single displayed colour in each municipality. It is worth remembering that positive correlations are indicated by colours ranging from green (both variables have low values) to violet (both high), while colours such as red and blue indicate negative correlations (one variable is high, and the other is low or vice-versa).

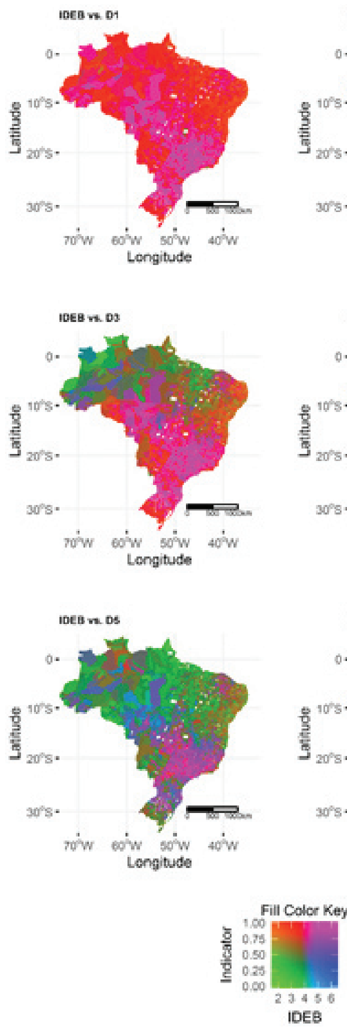

Figure 4. Choropleth maps of the values of IDEB and the dimensions composing IBEU as well as the *Socioeconomic Status* (SE) (F) for each municipality. The lack of data coverage on SE is visible in the bottom right map.

As expected from the discussion above, the choropleth map for the relation *IDEB vs Urban Mobility (D1)* (top left in Figure 4) is also the one that displays more reddish- and bluish-coloured (strong negative correlations) regions. This last point is compatible with the above-mentioned fact that the *D*1 dimension was observed as the only dimension that has a noticeable minor relation with the others of IBEU (de Queiroz Ribeiro & Ribeiro, 2013, pp. 29–30). This is even more understandable as the work-to-home travel time is even less relevant to the urban welfare in municipalities smaller than the metropolises for which IBEU was conceived.

The best correlation at the national level was *IDEB vs Urban Environmental Conditions (D2)*, as seen from Table 5. Nevertheless, at municipalities' level, the best correlation was visibly to Urban Infrastructure (D5) (bottom left in Figure 4), with less colour-mixing and mostly greenish- and a few violetish-coloured regions, and a little less so to Attendance of *Urban Collective Services* (D4) (middle right in Figure 4). This result suggests that these last two dimensions reveal better the spatial nature of and the distributional disparities in infrastructure access and ultimately the notion of urban welfare at municipalities' level. This seems to corroborate our suspicion of a relation between residential segregation and distributional disparities in infrastructure access and school inequalities.

Similarly, we built choropleth maps for the relations between IDEB and the top six indicators from Table 9 values for each municipality (Figure 5).

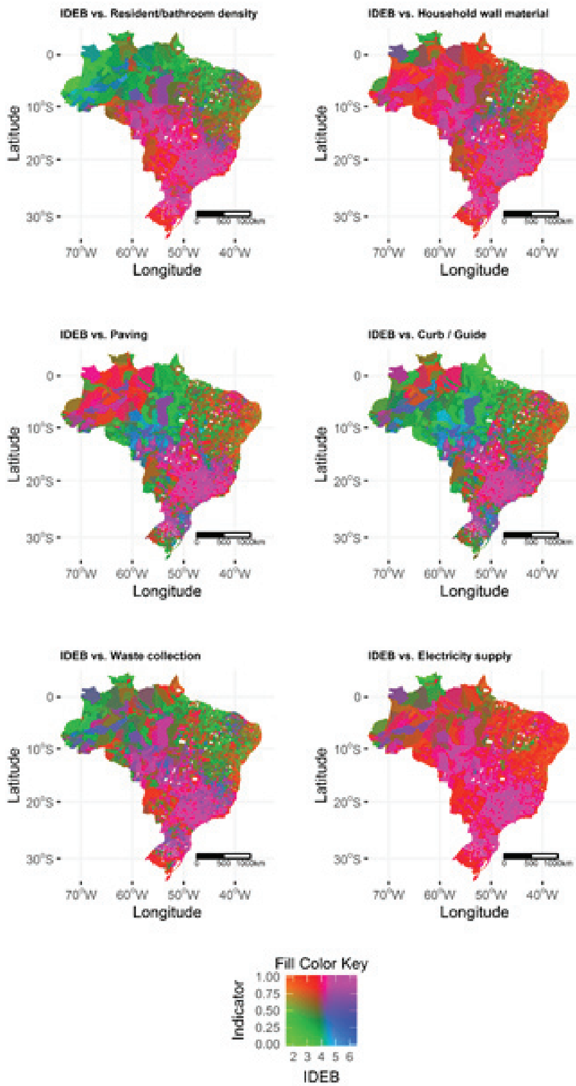

Figure 5. Choropleth maps of the values of IDEB and the top six indicators from Table 9 for each municipality.

The best individual indicator correlation at the national level was *IDEB vs Resident/bathroom density (D3)* (Table 8), which was confirmed at municipalities' level (top left in Figure 4), followed by to Household wall material *(D3)* (top right in Figure 4).

Our results also seem to confirm that even discounting for the of socioeconomic status effect the presence of the "neighbourhood effect" (de Queiroz Ribeiro, 2005; Nieuwenhuis et al., 2013), in the sense that the uneven distribution of urban conditions around students' households reduces educational opportunities and engenders social inequalities that are incompatible with a sustainable society.

Finally, our results also suggest that some adjustment is needed for IBEU, as this index seems not to work as well as a measure of urban welfare for the Brazilian municipalities as it does for metropolises.

CONCLUSIONS

Differently from the experiences of other countries during the first industrialisation, the fast-changing Brazilian urban reality segregates people in socio-spatial terms, distributes urban public resources unfairly and consequently threatens student's access to the structure of educational opportunities.

Two centuries after the first choropleth map was built in Nineteenth-Century by Charles Dupin, a civil engineer, senator, and member of the French Academy of Sciences, as a research tool for visualizing and generalizing statistical data (Korycka-Skorupa & Pasławski, 2017), its use as a research tool was unfortunately not yet sufficiently explored in the scientific literature (Dewandaru, Supriana, & Akbar, 2018). In this work, choropleth maps were invaluable as tools to analyse spatial inequalities visually across all the individual municipalities.

Our results indicate that the distributional inequalities in infrastructure access reduces educational opportunities, what will ultimately engender and perpetuate social inequalities, what is incompatible with the ideal of a sustainable society.

ACKNOWLEDGEMENTS

The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/CAPES (Higher Education Personnel Coordination) for the scholarship granted. The authors would also like to thank Dr Rossano André Dal-Farra from the Universidade Luterana do Brasil, for many enlightening discussions and suggestions.

AUTHORS' CONTRIBUTIONS STATEMENTS

R.P.D.S. supervised the project. R.P.D.S., M.Ş.B., and I.L.L. conceived the idea presented, discussed the methodology, and organized the theoretical discussion. I.L.L. collected the data. R.P.D.S. analysed the data. All authors discussed the results and contributed to the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The interested reader can access the data files, the details of this analysis, and/or the R code used from our GitHub repository: https://github.com/RenatoPdosSantos/urban-welfare-and-school-inequalities

REFERENCES

Andrade, E., & Soida, I. (2015). A qualidade do ranking das escolas de ensino médio baseado no ENEM é questionável. *Estudos Econômicos (São Paulo)*, *45*(2), 253–286. https://doi.org/10.1590/0101-4161201545221eai

Boudon, R. (1979). *L'inégalité des chances: la mobilité sociale dans les sociétés industrielles*. Paris: Armand Colin.

Brasil. INEP. (2015). *Nota técnica ENEM 2015 por escola - Indicador de nível socioeconômico (Inse) das escolas*. Brasília: INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Retrieved from http://www.sed.sc.gov.br/documentos/censo-278/indicadores-educacionais-1/enem/4742-nota-tecnica-indicador-nivel-socioeconomico/file

Brasil. INEP. (2016). *Divulgação ENEM 2015 por escola*. Brasília: INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Retrieved from http://download.inep.gov.br/educacao_basica/enem/enem_por_escola/2015/apresentacao_enem_por_escola_2015.pdf

Cameron, S. J. (2017). Urban inequality, social exclusion and schooling in Dhaka, Bangladesh. *Compare: A Journal of Comparative and International Education*, *47*(4), 580–597. https://doi.org/10.1080/03057925.2016.1259555

Carter, P. L. (2016). Educational Equality Is a Multifaceted Issue: Why We Must Understand the School's Sociocultural Context for Student Achievement. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(5), 142–163. Retrieved from https://www.rsfjournal.org/doi/full/10.7758/RSF.2016.2.5.07

Castells-Quintana, D. (2018). Beyond Kuznets: Inequality and the size and distribution of cities. *Journal of Regional Science*, *58*(3), 564–580. https://doi.org/10.1111/jors.12368

Coleman, J. S. (1966). *Equality of Educational Opportunity (Coleman) Study (EEOS), 1966*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR06389.v3

de Bruin, A., & Liu, N. (2019). The urbanization-household gender inequality nexus: Evidence from time allocation in China. *China Economic Review*, 101301. https://doi.org/10.1016/j.chieco.2019.05.001

de Queiroz Ribeiro, L. C. (2005). Segregação residencial e segmentação social: o "efeito vizinhança" na reprodução da pobreza nas metrópoles brasileiras. *Cadernos Metrópole*, (13), 47–70. Retrieved from http://www.redalyc.org/html/4028/402837794002/

de Queiroz Ribeiro, L. C., Koslinski, M. C., Zuccarelli, C., & Christovão, A. C. (2016). Desafios Urbanos à Democratização do Acesso às Oportunidades Educacionais nas Metrópoles Brasileiras. *Educação & Sociedade*, *37*(134), 171–193. https://doi.org/10.1590/ES0101-73302016157093

de Queiroz Ribeiro, L. C., & Ribeiro, M. G. (2013). *IBEU - Indice de Bem-Estar Urbano*. Rio de Janeiro: Letra Capital. Retrieved from http://www.observatoriodasmetropoles.net/images/abook_file/ibeu_livro.pdf

Dewandaru, A., Supriana, S. I., & Akbar, S. (2018). Event-Oriented Map Extraction From Web News Portal: Binary Map Case Study on Diphteria Outbreak and Flood in Jakarta. In *Proceedings of ICAICTA 2018 - 5th International Conference on Advanced Informatics: Concepts Theory and Applications, Krabi, Thailand, 14-17 Aug. 2018* (pp. 72–77). New York: IEEE. https://doi.org/10.1109/ICAICTA.2018.8541345

Downey, D. B., & Condron, D. J. (2016). Fifty Years since the Coleman Report: Rethinking the Relationship between Schools and Inequality. *Sociology of Education*, *89*(3), 207–220. https://doi.org/10.1177/0038040716651676

Fay, M., & Straub, S. (2017). *Rising incomes and inequality of access to infrastructure among Latin American households*. *Policy Research Working Paper 7987*. Washington, DC: The World Bank. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/26185/WPS7987.pdf?sequence=1

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 87. https://doi.org/10.1609/aimag.v17i3.1230

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3rd ed.). Los Angeles: Sage.

Friedman, E. (2018). Just-in-Time Urbanization? Managing Migration, Citizenship, and Schooling in the Chinese City. *Critical Sociology*, *44*(3), 503–518. https://doi.org/10.1177/0896920517695867

Harvey, D. (1971). *Social Justice and the City*. London: Edward Arnold. Retrieved from https://s3.amazonaws.com/arena-attachments/773624%252Fb9da5065225b2af72ffd9a15a26a5073.pdf

Hope, A. C. A. (1968). A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistical Society: Series B (Methodological)*, *30*(3), 582–598. https://doi.org/10.1111/j.2517-6161.1968.tb00759.x

IBGE. (2012). Censo Demográfico 2010. Retrieved June 28, 2017, from https://sidra.ibge.gov.br/pesquisa/censo-demografico/demografico-2010/inicial

Jasper, C., Le, T.-T., & Bartram, J. (2012). Water and Sanitation in Schools: A Systematic Review of the Health and Educational Outcomes. *International Journal of Environmental Research and Public Health*, *9*(8), 2772–2787. https://doi.org/10.3390/ijerph9082772

Kalogirou, S. (2012). Testing local versions of correlation coefficients. *Jahrbuch Für Regionalwissenschaft*, *32*(1), 45–61. https://doi.org/10.1007/s10037-011-0061-y

Kalogirou, S. (2013). Testing geographically weighted multicollinearity diagnostics. In *Proceedings of GISRUK 2013, Liverpool, UK, 3-5 April 2013*. Liverpool: Department of

Geography and Planning, School of Environmental Sciences, University of Liverpool. Retrieved from http://gisc.gr/?mdocs-file=1140

Kalogirou, S. (2015). A spatially varying relationship between the proportion of foreign citizens and income at local authorities in Greece. In *Proceedings of the 10th International Congress of the Hellenic Geographical Society, Thessaloniki 22-24 October 2014*. Thessaloniki: Aristotle University of Thessaloniki. Retrieved from htpps://gisc.gr/?mdocs-file=1048

Kalogirou, S. (2019, April 23). Package ‘lctools.’ Retrieved from https://cran.r-project.org/web/packages/lctools/index.html

Kim, S. (2015a). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665

Kim, S. (2015b, December 3). Package ‘ppcor.’ Retrieved from https://cran.r-project.org/package=ppcor

Korycka-Skorupa, J., & Pasławski, J. (2017). The beginnings of the choropleth presentation. *Polish Cartographical Review*, *49*(4), 187–198. https://doi.org/10.1515/pcr-2017-0012

Lewis-McCoy, R. L. (2014). *Inequality in the Promised Land: Race, Resources, and Suburban Schooling*. Stanford, CA: Stanford University Press.

Mack, W. (2017). Education in the City from the Perspective of Social Spaces. In *Education, Space and Urban Planning* (pp. 205–211). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-38999-8_19

Marqués Perales, I. (2016). Apuntes sobre el Informe Coleman. Sobre la Difícil Convivencia de los Principios Igualitarios en un Mundo Desigual. *International Journal of Sociology of Education*, *5*(2), 107. https://doi.org/10.17583/rise.2016.2092

Martín-Lagos López, M. D. (2018). Education and inequality: a meta-synthesis after the 50th anniversary of Coleman’s report. *Revista de Educación*, *380*, 176–198. https://doi.org/10.4438/1988-592X-RE-2017-380-377

Moran, P. A. P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, *10*(2), 243–251. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x

Murphy, W. (2016, November 1). colorplaner: A “ggplot2” Extension to Visualize Two Variables per Color Aesthetic Through Color Space Projections. Retrieved from https://cran.r-project.org/package=colorplaner

Nieuwenhuis, J., Hooimeijer, P., van Dorsselaer, S., & Vollebergh, W. (2013). Neighbourhood Effects on School Achievement: The Mediating Effect of Parenting and Problematic Behaviour? *Environment and Planning A: Economy and Space*, *45*(9), 2135–2153. https://doi.org/10.1068/a45367

Nieuwenhuis, J., Tammaru, T., van Ham, M., Hedman, L., & Manley, D. (2019). Does segregation reduce socio-spatial mobility? Evidence from four European countries with different inequality and segregation contexts. *Urban Studies*. https://doi.org/10.1177/0042098018807628

Peng, R. D. (2016). *Exploratory Data Analysis with R*. Victoria, CA-BC: Leanpub.

Permanyer, I. (2013). Using Census Data to Explore the Spatial Distribution of Human Development. *World Development*, *46*, 1–13. https://doi.org/10.1016/j.worlddev.2012.11.015

Pinheiro, D. R., Guardabassio, E. V., Bonjardim, E. C., & Bresciani, L. P. (2016). O desenvolvimento das metrópoles brasileiras segundo o IDH-M e o IBEU. *Espacios*, *37*(32).

PNUD, IPEA, & FJP. (2014). *Human Development Atlas in the Brazilian Metropolitan Regions 2013*. Brasília: PNUD Brasil. Retrieved from http://atlasbrasil.org.br/2013/en/download/publication/

R Core Team. (2019). The R Project for Statistical Computing. Retrieved March 11, 2019, from https://www.r-project.org/

Rey, S. J., & Smith, R. J. (2013). A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences*, *6*(2), 55–70. https://doi.org/10.1007/s12076-012-0086-z

Ribeiro, M. G. (2016). Desigualdades urbanas e desigualdades sociais nas metrópoles brasileiras. *Sociologias*, *18*(42), 198–230. https://doi.org/10.1590/15174522-018004207

Soares, J. F. (2004). The impact of the school in the cognitive achievement of the students (In Portuguese). *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educación*, *2*(2), 83–104. Retrieved from https://revistas.uam.es/index.php/reice/article/view/5550

Straub, S. (2011). Infrastructure and development: A critical appraisal of the macro-level Literature. *Journal of Development Studies*, *47*(5), 683–708. https://doi.org/10.1080/00220388.2010.509785

Tan, S. C., Ho, C. M., & Pang, V. (2015). *Education inequality: become better or worse? MPRA Paper*. München.

Umar, H. M. (2017). The Microeconomic Determinants of Educational Inequality in Nigeria. In S. F. Saidin, Z. M. Daud, A. Abdullah, & N. H. A. Razak (Eds.), *Proceedings of the Global Conference on Business and Economics Research (GCBER), Serdang, 14-15 Aug 2017* (pp. 345–350). Serdang: Faculty of Economics and Management, Universiti Putra Malaysia. Retrieved from http://www.econ.upm.edu.my/research/global_conference_on_business_and_economics_research_2017/gcber_2017_proceedings-32931

United Nations Development Programme (UNDP). (1990). *Human Development Report 1990*. New York: OUP - Oxford University Press. Retrieved from http://hdr.undp.org/en/reports/global/hdr1990

Vetter, D. M., Pinto, D. M. A., Fredrich, O. B. de L., & Massena, R. M. R. (1981). A proposição dos benefícios das ações do Estado em áreas urbanas: seus determinantes e análise através de ecologia fatorial. *Revista Brasileira de Geografia*, *43*(4), 455–476. Retrieved from https://biblioteca.ibge.gov.br/visualizacao/periodicos/115/rbg_1981_v43_n4.pdf

Wei, Y. D., Xiao, W., Simon, C. A., Liu, B., & Ni, Y. (2018). Neighborhood, race and educational inequality. *Cities*, *73*, 1–13. https://doi.org/10.1016/j.cities.2017.09.013

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer New York. Retrieved from https://www.springer.com/gp/book/9783319242750
Wickham, H., & Chang, W. (2019, August 11). Package 'ggplot2.' Retrieved from https://cran.r-project.org/web/packages/ggplot2/index.html