Society for the Study of Economic Inequality

Working Paper Series

Spatial divisions of poverty and wealth: How much does segregation matter for educational achievement?

Gabriel Otero
Rafael Carranza
Dante Contreras

ECINEQ WP 2020 - 543

ECINEQ 2020 - 543
June 2020

www.ecineq.org

Spatial divisions of poverty and wealth: How much does segregation matter for educational achievement? ‡

Gabriel Otero
Utrecht University

Rafael Carranza
London School of Economics

Dante Contreras
Universidad de Chile

Abstract

We explore how different spatial compositions affect the educational achievement in mathematics of 16 year-old students in Chile, a Latin American country with high income inequality and school segregation. We develop a critical review on the literature on negative ”neighbourhood effects” associated with concentrated poverty, complementing it with studies concerning self-segregation preferences by members of the upper-middle class. We combine administrative data about student performance with survey data for the 52 municipalities of the Metropolitan Region of Santiago de Chile. We cluster the districts based on factors such as unemployment, economic inequality, access to services, experiences of violence and stigmatization. Using longitudinal data, we look at the effect of each of the six spatial clusters on academic performance. Spatial clusters report a significant effect, above and beyond that of individual, household, and school-level characteristics. We conclude that space complements and reinforces the processes of accumulation of socioeconomic (dis)advantages.

Keywords: Class; Education; Inequality; Socio-economic environment; Urban studies.

JEL Classification: D63, I24, R23.

‡We are grateful for the support of the Centre for Social Conflict and Cohesion Studies (ANID/FONDAP/15130009).

Introduction

Spatial conditions have gained importance in scientific explanations of people's life opportunities above and beyond individual characteristics. Scholars have noted the impact of "neighbourhood effects" on social and personal outcomes such as child development, earnings, and well-being (Wilson, 1987; Jencks and Mayer, 1990; van Ham et al., 2012; Sharkey and Faber, 2014; Musterd et al., 2019; Sampson, 2019). A large part of this research agenda has focused on the negative effect of living in disadvantaged areas in terms of educational outcomes such as school dropout, level of instruction, and standardized test performance (Nieuwenhuis and Hooimeijer, 2016). Thus far, findings have been inconclusive, and research has been challenged due to a number of issues.

First, most studies have focused on determining how neighbourhood poverty relates to the educational outcomes of young people. Researchers have tried to measure the effect of neighbourhood socioeconomic status using variables such as poverty level, unemployment rate, criminality, and a lack of access to public services. However, some scholars have suggested that the spatial concentration of socioeconomic advantage is likely to be a stronger contributor to educational inequality than the concentration of disadvantage (Johnson, 2013; Howell, 2019a). Despite this possibility, little is known about whether living in affluent areas has a positive effect on students' academic performance (c.f., Ensminger et al., 1996; Gordon and Monastiriotis, 2006; Howell, 2019b). To shed light on this problem, this study examines the effects of living in different areas of the spatial context.

Second, academics have proposed a number of theoretical mechanisms for understanding how the spatial context can affect people's opportunities in life, such as exposure to violence, residential stigmatisation, and collective socialisation (see Galster, 2012). However, how these "spatial forces" operate to affect specific social outcomes is still unclear (Harding, 2011; Sharkey and Faber, 2014). This analytic problem has affected the literature partly due to three closely related empirical deficits: (1) the spatial mechanisms have rarely been combined in research in order to understand their intersectionalities; (2) few studies have explicitly examined their bidirectionality or asymmetrical forms of causality (c.f., Brooks-Gunn et al., 1993; Johnson, 2013; Howell, 2019b); (3) the thresholds of their effects have not been studied either (Musterd et al., 2019). To help alleviate these issues, this study enriches the conceptualization of advantaged neighbourhoods and delineates how they may facilitate the intergenerational transmission of socioeconomic advantages. In empirical terms, we focus on representing the spatial structure, that is, the various spatial arrangements observable depending on a number of characteristics (e.g., income levels, economic inequality, access to services). Then, we determine the distinctive effect of these arrangements on academic performance and establish whether they increase a student's chances to score in the top or bottom 25% of all students taking a standardized test.

Third, several studies have focused on testing how spatial characteristics influence educational outcomes after considering students' socioeconomic background (i.e., parents' education or household income). Nevertheless, from an ecological perspective, the school context should also play a relevant role in explanations of school learning, especially given the multiple resources offered by institutions as well as the decisions and practices implemented there (Bronfenbrenner, 1979). The joint analysis of these contexts has demonstrated to be relevant in empirical terms. First, including factors related to school composition may reduce omitted variable bias (van Ham et al., 2012). Second, the demographic composition of neighbourhoods and schools certainly overlaps to a large extent; therefore, the effects of both contexts are likely to be interdependent (Nieuwenhuis et al., 2016). The handful of studies that have included variables to represent the school context have reported mixed evidence, tending to find weaker "neighbourhood effects" (e.g., Kauppinen, 2008; Owens, 2010; Sykes and Musterd, 2011; Otero et al., 2017). To address these issues, we separately test the relevance of the contextual effects of residential areas, school, and household/family on academic achievement.

The focus of our inquiry is on Chile, an upper-middle income Latin American country with one of the most unequal income distributions, and one of the most segregated educational systems in the world – which has increased in recent years (Valenzuela, et al. 2014). We specifically analyse the Metropolitan Region of Santiago, home to nearly 40% of the population. We intend to find answers to the following research questions:

(1) Do spatial characteristics have a significant effect on the academic performance of Chilean adolescents, above and beyond family background, school context, and students' individual characteristics?
(2) If significant "neighbourhood effects" are found, what are the specific spatial compositions that actually result in improved or worsened academic performance?
(3) To what extent are the effects of the spatial context weakened when other social contexts are simultaneously considered?

To answer these questions, we used a sample of 37,000 adolescent 10th graders (aged 15-16). We combined a panel of administrative data about student performance on a standardized mathematics test with survey data for the 52 municipalities of the Metropolitan Region of Santiago de Chile.[2] In the next section, we examine theoretical arguments about the influence of spatial segregation on educational

[2] We focused on the Mathematics test for three reasons: (1) fewer students take the Language test than the Mathematics one, (2) the explanatory power of the model is significantly higher for Mathematics, and (3) Mathematics performance is a better proxy for future income (Rose and Betts, 2006).

achievement, making specific points regarding the potential effects of marginalisation and concentrated affluence.

Theoretical background

Spatial segregation and educational inequality

In this study, we adopt a multidimensional framework for understanding the ways in which multiple social processes (e.g., material, symbolic, cultural, and spatial) contribute to the production and reproduction of inequality (Lamont et al., 2014). Regarding spatial processes, we share the assumption that spatial inequality should reinforce the pre-existent socioeconomic inequality generated by macrosocial factors (Sampson, 2019). In our study, we specifically expect spatial conditions to have an independent effect on academic performance after considering aspects such as the educational system and wealth distribution.

Spatial segregation by socioeconomic status may well be the key mechanism for understanding differences in educational outcomes generated by the spatial context. The concept of segregation is generally used to refer to the extreme spatial concentration of social groups, that is, affluent neighbourhoods, gated communities, ghettos, or enclaves (Marcuse, 2005). It is assumed that spatial segregation is mainly produced by government and real estate market devices to impose and intensify patterns of urban exclusion based on race/class, as well as by the preferences of the higher classes, which have more residential choice opportunities (Bourdieu, 1993; Massey, 1996; Slater, 2013; Pinçon-Charlot and Pinçon, 2018).

Regarding the first condition, it should be noted that neighbourhood classification is not random; rather, it is strongly structured by the socioeconomic resources available to families (unequal opportunity structure) and housing availability (real estate market) (Gingrich and Ansell, 2014). In highly unequal contexts, such as Global South cities in general, spatial segregation by socioeconomic level is much more widely extended than in the welfare regimes typically found in Western Europe (Musterd et al., 2017). In Chile, for instance, neoliberal reforms in the domains of health care, education, and pensions have been introduced within the framework of a subsidiary state (Angotti, 2017). Market-led public housing policies have generated or deepened urban marginalization and exclusion through devices such as the allocation of social housing in the periphery of cities (Caldeira, 2017). The logic of political economy has restricted the role of the state in the establishment and assurance of sufficient living conditions and strategies for all residents, thus conditioning the opportunities available to the poor.

Regarding the second condition, privatisation processes –which have a long history in Chile– have facilitated the expansion of neoliberal subjectivity, leading to the generation and/or preservation of spatial segregation mainly based on residential choice (Méndez and Gayo, 2019). This ability to choose is logically much more likely for the middle (high) classes than for the poor, who are usually trapped in the most devalued areas of the city and have limited mobility opportunities (Slater, 2013). The residential choices of the (upper) middle classes, which cannot always be understood as strategic and/or "rational" actions, have tended to be based on a visible preference for living "among equals" (Andreotti et al., 2015); for instance, through processes such as urban expansion, the emergence of gated communities, and gentrification (i.e., colonization of neighbourhoods and exclusionary displacement by the wealthiest classes) (Bridge et al., 2012). The high prices of the homes preferred by the higher classes in affluent districts tend to reinforce the physical barriers separating the richer population from the rest and enable wealthier groups to control and shape places (Maloutas and Pantelidou Malouta, 2004). In this context, a dual type of exclusion can be said to affect people with scarce capital: spatial fixation in vulnerable areas and limited possibilities of urban accessibility (Bourdieu, 1993).

By highlighting the restrictions experienced by people living in low-income or poor areas, compared to the choices available to more affluent middle-class people, we argue that inequality in terms of socio-spatial opportunities is a relational problem that must be analysed considering the absolutely fundamental structural conditions –especially the privileges enjoyed by those who occupy the higher positions within the class and spatial structures (Massey, 1996; Pinçon-Charlot and Pinçon, 2018). This issue has been noticeably overlooked in the literature on "neighbourhood effects" and educational outcomes, which has almost exclusively focused on the negative effects of living in underprivileged areas (Howell, 2019b). By leaving out the topmost part of the spatial structure, researchers have limited their ability to understand which mechanisms matter in various areas, not only impoverished ones (Johnson, 2013). In the following sections, we elaborate on how spatial segregation, expressed through concentrated poverty and wealth, could reinforce educational inequality at both ends.

The mainstream mechanisms of neighbourhood effects on educational outcomes

Researchers of "neighbourhood effects" have suggested that neighbourhood poverty (or concentrated disadvantage) have a negative influence on educational outcomes mainly through four mechanisms: social isolation, social disorder, lack of institutional resources, and environmental hazards.

In general, social isolation theories suggest that residents of impoverished urban areas are quite disconnected from the social networks of the middle classes and traditional institutions (Wilson, 1987). This strong social isolation is believed to generate major obstacles in the socialization processes of children and adolescents, making it difficult to stimulate academic success at school and convey the advantages of formal education (Jencks and Mayer, 1990; Leventhal and Brooks-Gunn, 2000; Ainsworth, 2010). Some scholars have specifically noted that the concentration of disadvantages may have a negative impact on the educational outcomes of adolescents due to the presence of heterogeneous (competitive and conflictive) cultural models of schooling –which include both "middle class" dispositions and orientations and lifestyles associated with "alternative" cultures (Harding, 2007). In this context, it is presumed that students have fewer chances to act consistently with the aims that they articulate regarding their educational and occupational future; also, they may find it easier to change their path if their social environment provides them with other choices (Harding, 2011).

Explanations associated with social disorganization have focused on the internal processes of neighbourhoods, mainly on social cohesion attributes. It is presumed that residents of underprivileged neighbourhoods encounter difficulties that prevent them from maintaining social ties and support networks and joining social organizations (Wacquant, 2008). In consequence, these residential settings display a reduction in their ability to supervise their members' behaviour and implement informal social control measures, especially regarding youth behaviour (Sampson and Groves, 1989; Sampson et al., 1997). Lacking "collective efficacy", children and adolescents from underprivileged neighbourhoods are more likely to develop behaviours that interrupt their educational progress (Leventhal and Brooks-Gunn, 2000; Morenoff et al., 2001).

Scholars who have examined institutional resources suggest that "neighbourhood effects" operate indirectly through the quality of the services available in residential areas. It is assumed that low-income or poor neighbourhoods frequently lack supermarkets, green areas, and cultural venues, but it is the absence of health care services and quality schools that makes it especially difficult to encourage children's development (Wilson, 1987; Small and Newman, 2001). Specifically, educational institutions located in marginalized residential environments often encounter serious budgetary difficulties to acquire teaching resources such as textbooks, highly qualified teachers, and support programs. In this context, students' educational aspirations and skills for achieving their academic goals are truncated or significantly affected (Bramley and Karley, 2007; Ainsworth, 2010; Lareau, 2011).

Environmental theories of neighbourhood have focused on area-based inequality in the exposure to air and water pollution, ambient noise, lead, and other contaminants (Diez Roux and Mair, 2010). These problems may be especially serious for children and adolescents who live in the most impoverished areas of cities, which tend to be located near motorways, waste sites, and manufacturing facilities (Crowder and Downey, 2010). Students living in these disadvantaged residential settings also tend to be overexposed to houses in a state of disrepair, which can severely affect their health due to the presence of toxins. These environmental hazards lead to health problems (e.g., respiratory disease, allergies, mental stress) and school absenteeism, which directly affect students' cognitive skills and achievement (e.g., Chen et al., 2018; Shier et al., 2019).

Some academics have stressed the importance of focusing on social dimensions of those environmental constraints, especially young people's exposure to violence. It has been suggested that both direct and indirect exposure to violent situations and criminal environments, such as witnessing a shooting or stabbing or being threatened with a weapon (Harding, 2009), may be particularly damaging to cognitive functioning and academic performance (e.g., Sharkey, 2010; Caudillo and Torche, 2014). The specific mechanisms leading to these outcomes likely involve symptoms of psychological post-traumatic stress, shock, experiences of fear, anxiety, depression, and speech impediments (Sharkey and Faber, 2014).

Complementary mechanisms that have received attention from researchers include spatial stigmatisation and relative deprivation. Regarding the "blemish of places", scholars have suggested that more qualified teachers may not want to work in schools located in stigmatized areas, which can lead young people there to underestimate educational achievement (Galster et al., 2007). With respect to the second mechanism, it has been suggested that visible economic inequalities within a residential environment may have a negative impact on educational outcomes, especially when extreme inequalities are observed (Otero et al., 2017).

Understanding the "privilege" of those living in wealthy areas for explaining better educational outcomes

As we have argued, the traditional focus of the "neighbourhood effects" literature on concentrated poverty has overlooked the fact that the exclusion of the most marginalized can be partly due to the segregation of the most affluent. As we have argued, the traditional focus of the "neighbourhood effects" literature on concentrated poverty has overlooked that the exclusion of the most marginalized can be partly due to the segregation of the most affluent. At the macrosocial level, the high price of homes in affluent areas of the city, where high performing schools are usually located, can enable richer people to "target" access to (higher quality) education in a way that excludes low-income citizens (Gingrich and Ansell, 2014). In the case of highly stratified (and neoliberal) educational systems, the main mechanism that explains the strong link between residential segregation and school segregation is the concomitant selection between parents and schools (Méndez and Gayo, 2019). Here, the educational system is highly privatised and accordingly, the quality of education is highly dependent on tuition fees. As such, high performing schools are usually located in affluent areas because in those areas parents are capable to overcome the strong financial barriers to entry to private schools.

At the microsocial level, the choices available to the upper echelons and their preferences play a major role in reinforcing segregation patterns. This is precisely the point made by scholars concerned with the schooling and residential strategies of the upper middle classes in several cities of Latin America and Europe (e.g., Bridge et al., 2012; Andreotti et al., 2015; Bacqué et al., 2015; Méndez and Gayo, 2019). They have demonstrated that the upper-middle classes operate as social "segregators"; that is, as a group that aspires to live "among equals" in order to accumulate cultural and social capital and facilitate upwards social mobility for their children, among other goals.

This preference of the (upper) middle classes for living among "peers" in affluent residential environments has major implications for young people's socialization. When it comes to educational aims, affluent residents are expected to be much more homogeneous than the rest and to orient themselves toward the expectation of attending university (Harding, 2011). This homogeneity in their expectations may be explained by the type of parenting that commonly characterises this privileged group, usually referred to as "concerted cultivation" (Lareau, 2011) – which is focused on developing children's talents and skills by encouraging extracurricular cultural activities in their free time. These practices are aimed at fostering children's and adolescents' motivation and confidence, two attributes that can be especially useful when addressing school demands (Bourdieu and Passeron, 1977). Furthermore, this shared cultural orientation towards education is generally reinforced through collective socialization in areas characterized by the concentration of affluence (Lareau and Goyette, 2014). Therefore, this can result in strong normative pressures on privileged adolescents to attain better academic performance.

Several sociological accounts have suggested that the preference of the higher classes for socializing with their equals in segregated residential environments provides substantial advantages for cementing stronger neighbourhood cohesion, i.e., residential belonging, class identity, and performative practices of place-making (e.g., Savage et al., 2005; Andreotti et al., 2015). This is not a type of "collective efficacy" to tackle crime and antisocial behaviours, but rather a set of attitudes focused on working for the neighbourhood and improving residential conditions in a broader sense (Méndez and Gayo, 2019). As such, social organization and sociability may be understood as an instrument that residents of affluent (segregated) neighbourhoods can use as a fundamental part of their privilege (Maloutas and Pantelidou Malouta, 2004; Méndez et al., 2020). When applied to education, social cohesion is expected to increase the ability of adult residents of affluent neighbourhoods to instil a consistent set of positive norms and values regarding education in young people, such as the advantages of achieving good educational outcomes (Coleman, 1988). These attributes have often been linked to academic achievement in empirical research (e.g., Woolley and Grogan-Kaylor, 2006; Ainsworth, 2010).

Apart from the well-known preference of the upper middle class for living in more homogeneous and safe environments in order to benefit from a specific symbolic value, it is also well established that this group strives to access high-quality institutions through its residential choices (e.g., Butler and Hamnett, 2007; Bacqué et al., 2015). Therefore, wealthy neighbourhoods typically have more green spaces, public recreational sport facilities, cultural infrastructure (e.g. museums, theatres, libraries), and better schools (e.g. good infrastructure, highly educated teachers, reinforcement lessons). This combination of symbolic and pragmatic advantages allow upper middle class residents to support their class sociability in a variety of social settings, thus increasing the likelihood of interacting "among equals" and restricting contact with dissimilar others (Lareau and Goyette, 2014). Finally, it is highly probable for the available resources and facilities to result in better child development through better healthcare provision, while also supporting young people's educational activities and providing them with the support necessary to attain educational success (Gingrich and Ansell, 2014; Johnson, 2013; Leventhal and Brooks-Gunn, 2000).

Hypotheses

An assumption shared by most researchers is that the spatial context is an additional source of advantage/disadvantage and that it can therefore play a major role in reinforcing economic inequalities. Considering this, we expect to identify effects of the spatial context on academic performance other than those associated with individual, home, and school characteristics. This prompts our first hypothesis:

Hypothesis 1: The spatial context will produce significant differences in academic performance in mathematics, independent of the effect associated with other contexts (household, school).

The emphasis the "neighbourhood effects" literature on the negative consequences of living in impoverished areas has implicitly suggested the idea that schools with limited resources are especially ineffective and that they are ultimately responsible for their own (poor) socioeconomic situation and the weak academic performance of the children and adolescents that they serve. However, the consequences of spatial segregation on various social outcomes are highly likely to be related not only to the exclusion of the working classes and poverty concentration, but rather to residential choice opportunities and, more specifically, to the preference of the wealthier classes for living near similarly advantaged peers. In consequence, "neighbourhood effects" should be much more powerful and impactful at the top of the spatial hierarchy than at the bottom. One way of evaluating this expectation is to compare gaps in concentrated advantage/disadvantage with respect to intermediate positions. With this in mind, we propose an additional hypothesis:

Hypothesis 2: Concentrated advantage will produce a larger academic performance gap than concentrated disadvantage or poverty, compared to middle SES areas.

Lastly, analysing the effects of the spatial context on academic performance logically involves considering the level of association between the various settings that can influence educational outcomes, i.e., family background and school context. This issue could be particularly relevant for the Chilean context. Since the educational reform of 1981 (during Pinochet's military dictatorship), Chile has developed an educational system driven by market policies, privatisation, and school choice (Carrasco and Gunter, 2019). There are three types of schools: public schools, serving approximately 35% of all students, subsidised schools (55%), and private schools (about 10%). Families with more economic capital can afford expensive tuition fees and enrol their children in fully private schools (Gayo et al., 2019) – where the highest scores on standardized tests of academic performance are usually attained (e.g., Contreras et al., 2010; Mizala and Torche, 2012). This has turned the Chilean educational system into one of the world's most segregated in socioeconomic terms (Valenzuela et al., 2014).

In addition, there is a clear association between the segregation of the educational system and spatial segregation, especially in Santiago, the capital of Chile (Santos and Elacqua, 2016). In general, the most impoverished areas only have access to public schools, which score lower on standardized tests (Otero et al., 2017). In contrast, more mixed and middle-class areas tend to have higher quality State-subsidised private schools, while affluent residential settings concentrate the majority of private schools. Certainly, attending private schools is a sign of status, and probably one of the reasons why social stratification prevents upward social mobility (Méndez and Gayo, 2019). Ultimately, access to these schools has a major influence on students' possibilities of entering a leading university and joining the country's most respected professionals (Zimmerman, 2019).

In brief, the clear socioeconomic segregation by socioeconomic level present in contexts with the potential of affecting educational outcomes suggests that the effects of the spatial context should be considerably reduced when other contexts are simultaneously included in the models. This prompts our last hypothesis:

Hypothesis 3: The effect of the spatial context on educational achievement will be smaller when taking into account family, school, and individual characteristics.

Data and methods

Data

This study uses panel data from a sample of 37,000 adolescent students who took the Mathematics SIMCE test in 4th grade (2010, 10 years old), again in 8th grade (2014, 14 years old), and a third time in 10th grade (2016, 16 years old). The SIMCE also includes information about test takers' households and the schools that they attend. Information is collected through a set of questionnaires completed by parents and school administrators. The data provided by the SIMCE were complemented with a survey covering the 52 municipalities of the Metropolitan Region of Santiago, taken from the National Socioeconomic Characterization Survey [Encuesta de Caracterización Socioeconómica Nacional, CASEN], carried out from November 2015 to early 2016.

We merged the district level estimates from the 2015 CASEN with the administrative data for each individual student using their school's district. We do not use their home's district as the SIMCE data does not identify them. We therefore focus on the role played by the school neighbourhood. As not many students change districts over their life as students, these effects are persistent over time. Our data shows that, for the Metropolitan Region, 65% of students have remained in the same district between 4th and 10th grade, and over 75% have stayed in the same district between 8th and 10th grade.

As not at test-takers have complete information, we lose observations in the process of arranging the data. Table 1 shows the number of observations in each step of the process. The SIMCE database includes almost 250 thousand students, but only 196 thousand actually took the exam. Among them, 162 also filled the parents' questionnaire, and 141 thousand included information on their parents' education and income. This leaves us with 57.2% of the original population of students. Furthermore, part of the test-takers ID numbers are duplicated and do not allow for an accurate merge across years.[3] After dropping the duplicates, we are left with 115 thousand cases, among which 101 thousand take all three SIMCE tests (2010, 2014, and 2016). The following step is to merge the SIMCE data for each student with their corresponding district-level data from CASEN. We have a small reduction in cases as a few small districts were not surveyed. As a result, we consider 41% of the original number of students. Finally, as we focus our study to the Metropolitan Region of Santiago, we are left with 37 thousand observations, which is equal to 15% of the overall population, or 36.4% of the feasible sample.

<<TABLE 1 ABOUT HERE>>

[3] To avoid losing all these observations, we also use gender and school as additional variables to refine the merge between datasets, i.e., we eliminate all duplicates with the same gender and in the same school.

Table 1 shows that the reduction in observations is not random. The average SIMCE score increases in each step, suggesting that lower-scoring students are more likely to be missing from the final sample. However, this is a small change, considering that one standard deviation equals 50 SIMCE points. The increase between the full sample and the 2015 CASEN sample amounts to 27.8% of a standard deviation. The last step increases the average even further, but this is product of our choice to focus on the Metropolitan Region of Santiago rather than an unavoidable loss of information. Although the average score increases, it is safe to say that the overall composition of our sample does not change substantially.

Variables

Dependent variables

Our variable of interest is the mathematics score of 10th year students (aged 15-16). From which we construct three dependent variables.

(1) 10th grade maths score
(2) Scoring within the top 25% (binary)
(3) Scoring within the bottom 25% (binary)

These three variables allow us to explain not only average performance, but its distribution across students.

Measuring the spatial context

We are interested in the spatial context as a factor for explaining differences in students' mathematics performance. To do so, we focus on the Metropolitan Region of Santiago. This region clearly reflects the extreme economic inequalities of Chile. It is the largest region in the country and concentrates about 40% of the population –some 7 million inhabitants. It is divided into 52 municipalities that greatly differ in terms of their residents' SES, service availability, and crime levels (Ruiz-Tagle and López, 2014; Garretón, 2017). The spatial segregation by SES has often described as the key defining characteristic of the Metropolitan region of Santiago, which is only comparable to spatial segregation by race found in some cities of the US (Agostini et al., 2016). We focused on the district scale, which is the one that most clearly represents the spatial segregation found in Santiago

We contribute to the literature by proposing a holistic characterization of students' surroundings. Instead of separately examining the effect of each spatial characteristic on academic performance, we aimed to represent spatial clusters of districts with shared characteristics. This approach is particularly useful for the identification of homogeneous areas, especially the spatial divisions of poverty and wealth underlined in our theoretical background. Specifically, we applied Hierarchical Clustering on Principal Components (HCPC). Here, different principal components methods are used as a pre-procedure in order to facilitate a more reliable clustering solution later (see Husson, Lê, & Pagès, 2011).

The first step of the HCPC is based on a Principal Component Analysis (PCA) to reduce the georeferenced (continuous) data to a smaller set of dimensions. We used 28 spatial characteristics including poverty rate, the proportion of residents with a complete higher education degree, income inequality, experiences of violence, accessibility to green spaces and services, among others (see Table A1 in the Appendix). In our analysis, the first two principal components explain 64.4% of the total variance, which is an acceptable portion (Abdi and Williams, 2010). The first component of the PCA represents the accessibility to services (or more broadly, the rural/urban divide), while the second component clearly represents the SES polarization across the region.

The second step of the HCPC was performed by means of a hierarchical clustering using Ward's criterion on the two selected principal components, and the last step was conducted using a k-means clustering algorithm to improve the previous partition. Based on these analyses, we arrived at a six-spatial cluster solution. Table 2 briefly describes the spatial clusters identified, and Section A (in the Appendix) presents more detailed information regarding the distribution of spatial characteristics between spatial clusters.

<<TABLE 2 ABOUT HERE>>

Figure 1 shows a factor map to represent how the 52 districts comprising the Metropolitan Region of Santiago are located with respect to the two principal components and their spatial clusters. The first component (dimension 1) makes a distinction between districts with low accessibility (clusters 2 and 3, to the left) and districts with medium-high accessibility to services (clusters 1, 4, 5, and 6, to the right). The second component (dimension 2) shows socioeconomic differences, with high income districts at the top (notably Cluster 6), and poorer districts at the bottom (notably Cluster 1). As shown in Figure 1, Cluster 6 is the most distinct one, being the farthest one from the rest of the clusters, and clearly represents the condition of living in affluent areas. In general, the spatial segmentation identified is visibly consistent with the SES-based spatial segregation patterns reported in previous research (e.g., Agostini et al., 2016; Garretón, 2017; Ruiz-Tagle and López, 2014).

<<FIGURE 1 ABOUT HERE>>

Control variables

To account for omitted variable bias and report the effects of the spatial context more accurately we include controls at three different levels: individual traits, household characteristics, and school environment. To represent the household context, we have included two measures: parents' years of schooling and monthly household income. The school context is represented by the type of school that each student attends. Individual level characteristics include gender and SIMCE scores in 4th grade and 8th grade. Past scores are particularly relevant since they are a central part of our causal identification strategy. Table 3 presents descriptive statistics for individual, household, and school variables.

<<TABLE 3 ABOUT HERE>>

Empirical strategy

Our empirical strategy addresses the biases arising from past outcomes. Ashenfelter and Card (1985) report how the decision to opt for a training program is heavily influenced by past earnings, as it is those with low earnings that opt for them. In a similar vein, past scores could affect whether students change schools or district – or even household characteristics such as family income – thus biasing the estimation for these coefficients. In other words, previous test scores are a time-varying confounder and we need to account for the longitudinal structure of the student's test scores.

We use a lagged dependent variable as our identification strategy (Angrist and Pischke, 2009). As our outcome is the Mathematics SIMCE test score in 10th grade, we control for previous tests scores for the same student, in 8th and 4th grade. Our specification for a given test score T for student i, in school j, and cluster k, in time period t:

$$T_{ijk,t} = \alpha + \beta X_{it} + \gamma S_{jt} + C_{kt} + \delta T_{i,t-1} + \eta T_{i,t-2} + \varepsilon_{ijkt}. \tag{1}$$

X_{it} are individual and household characteristics in time t, S_{jt} are school characteristics in time t, C_{kt} are district cluster fixed effects, $T_{i,t-1}$ and $T_{i,t-2}$ are past test scores, and ε_{ijkt} is an idiosyncratic error term. The outcome of interest is the Mathematics SIMCE score of the students in 10th grade, while the two past scores are from 8th and 4th grade, respectively.

We include two additional outcomes in order to study the symmetry of these effects, that is, whether cluster effects are simultaneously predicting 'good' and 'bad' outcomes. The two outcomes focus on specific parts of the SIMCE distribution: the former captures top scores, while the latter captures low scores. Following equation (1), we have the following model:

$$D^{a}_{ijk,t} = \alpha + \beta X_{it} + \gamma S_{jt} + C_{kt} + \delta T_{i,t-1} + \eta T_{i,t-2} + \varepsilon_{ijkt} \quad (2)$$

Where $a = \{top25, bottom25\}$.

The outcome of equation (2) is a binary variable. $D^{top25}_{ijk,t}$ takes the value one if the student's score lies between percentiles 75 and 100, while $D^{bottom25}_{ijk,t}$ takes the value one if the student's score lies between percentiles 1 and 25. In our sample, being in the top 25% requires a score above 332 points, while being in the bottom 25% requires a score below 240 points. The goal of these binary variables is to go beyond the linear effect between clusters and scores by testing whether the clusters have a differentiated effect on achieving high or low scores.

Equation (1) is estimated using OLS, while equation (2) is estimated using logit regression. In both cases, standard errors are clustered at the school level.[4]

Results

Descriptive analyses

Figure 2 reports the density plot for the 10th grade maths score for each cluster. We see the distinction between Cluster 6 (the high-income districts) and the rest. Cluster 6 shows a concentration of high scores with little dispersion. Clusters 3, 4, and 5 show similar distributions, each more skewed to the left than the former. Lastly, Cluster 1 and 2 show the bimodal distribution common in rural areas of Chile, with two separate groups of high- and low-test scores. The figure paints a picture of interrelated spatial and educational inequalities in the region, between clusters and in some cases, within clusters.

<<FIGURE 2 ABOUT HERE>>

[4] As an alternative to OLS, we tested multilevel estimations using the school and cluster and only cluster levels. Results do not vary substantially from those presented here.

Figure 3 reports the distribution of students in the top 25% of the 10th grade SIMCE test by cluster, conditional on the score that they obtained in the 4th grade SIMCE test (when they were 10 years old). The probability of being in the top 25% in 10th grade increases as 4th-grade scores increase, but these trends differ by cluster. The high-income cluster (Cluster 6) shows a higher share of students in the top 25% at all levels of the 4th grade score. A student with 250 points in 4th grade (the mean score) has a 10% to 20% chance of being in the top 25% in 10th grade for all clusters but cluster 6, where students have a 35% chance of being in the top 25%. The gap between Cluster 6 and the remaining clusters decreases for higher test scores, suggesting that good past performance can somewhat compensate for district effects. But even at the very top, that gap remains. Indeed, when we look at 350 points in 4th grade (the 99.7 percentile), the chance of being in the top 25% goes from 40% to 60% for clusters 1, 3, 4, and 5, while Cluster 6 shows a chance of almost 80%. Cluster 2 has an almost 70% chance at that level, with a change in slope consistent with its bimodal distribution.

<<FIGURE 3 ABOUT HERE>>

Regression analyses

Table 4 shows the result of all regressions. Models 1 to 4 are the OLS results for the average test score, while Model 5 is the logit regression for the probability of being in the top 25% and Model 6 is the logit regression for being in the bottom 25%. All logit coefficients are odds ratios, and therefore interpreted as the percentage change in the outcome variable when studying in a given Cluster. In all estimations our reference cluster is Cluster 3, a lower-middle-class cluster.

<<TABLE 4 ABOUT HERE>>

Model 1 shows the unconditional differences in test scores by spatial cluster. Relative to Cluster 3, all clusters show a statistically significant difference, with clusters 1 and 2 being just below Cluster 3, with 1.3 and 3.4 fewer points respectively. As expected, Cluster 6 shows the largest differences, having 58.8 points – 1.2 standard deviations – more than Cluster 3. Clusters 4 and 5 also show higher tests scores, with 11.8 and 20.6 points over Cluster 3, respectively.

Models 2, 3, and 4 gradually include additional control variables to explore the change in the cluster's coefficients. Model 2 includes the two lagged test scores in order to account for unobserved factors, following a lagged dependent variable identification strategy. As expected, previous scores strongly affect educational achievement in 10th grade, with the R-squared going from 7% in Model 1 to 58.2% in Model 2. Model 3 adds students' gender along with three variables accounting for the household context: each parent's years of schooling and monthly household income (in logs). All three household level variables show a positive and significant effect. Model 4 incorporates school characteristics, specifically the type of school in which the students are enrolled. Relative to being enrolled in a private-subsidised school, attending a public school has a negative effect (7.9 fewer points) and attending a private school has a positive effect (8.4 more points). These are expected results that match previous research on the influence of household and school contexts on academic achievement.

Going from Model 1 to Model 4 closes the gap between clusters, as the rest of the variables explain an important part of the performance gap. In Model 1, the gap between Cluster 6 and Cluster 1 is equal to 60.2 SIMCE points (or 1.2 standard deviations). In Model 6, the gap equals 3.7 points (or 0.07 standard deviations). As expected, the biggest change happens when including previous test scores in Model 2, which reduces the gap to 12.9 points (0.25 standard deviations). The gap more than halves when accounting for household characteristics in Model 3. Overall, individual, household, and school characteristics explain a large part, but not all of the gap between clusters.

Model 4 indicates that all clusters are different from Cluster 3 to a statistically significant degree, except for Cluster 1. Clusters 4, 5, and 6 –comprising middle-class, upper-middle-class, and affluent districts– show a positive difference (2.6, 2.8, and 4.6 points respectively). Conversely, Cluster 2 –the rural cluster with a bimodal socioeconomic distribution– shows a negative difference of 7.3 points. Our estimates show that the spatial context has a causal effect in different clusters, above and beyond the effect of personal, household, and school characteristics.

Even though the coefficients decrease once we account for other factors, the effects prove to be substantive. A 4.6-point increase –the difference between Cluster 6 and Cluster 3 in Model 4– is equivalent to a 2.6% increase in household income, or for a parent to have 5.5 more years of schooling. It is also equivalent to half of the effect of attending a private school (relative to being in a subsidised-private school). Our findings suggest that the effect of living in a certain area can be equivalent to the impact of well-known determinants of test scores.

We have seen that the effects of studying in different clusters go in predictable directions. However, the average effect could be hiding nonlinear effects. For example, an increase in the average score can be explained by clusters encouraging a higher performance or by discouraging a low performance (or both). We explore this question in Models 5 and 6 in Table 4, where the outcome is being in the top 25% and bottom 25% of test scores, respectively.

When compared to Cluster 3, all clusters but Cluster 1 have a significant effect on the probability of being in the top 25%. Cluster 2, 4, 5, and 6 increase their chance of being on the top 25%. On the other hand, all clusters reduce their chance of being in the bottom 25% relative to Cluster 3, except for Cluster 2. Overall, we see that high-income clusters (4, 5, and 6) both increase their chances of high achievement and reduce their chances of low achievement – particularly cluster 6. Cluster 1 only reduces their chances of low achievement (thus explaining the average increase in columns 1 to 4). Cluster 2 increases the chances of being in the top 25% and in the bottom 25% of test scores, but the latter is larger than the former, explaining the average decrease shown in columns 1 to 4. The strongest effects are associated with Cluster 2, which displays a 21.4% increase in the probability of being in the top 25%, and Cluster 6, which shows both a 17.7% increase in the probability of being in the top 25% and a 35% decrease in the probability of being in the bottom 25%.

With respect to the remaining variables, almost all show the dual effect of increasing the chance of high scores and reducing the chance of low scores, with the exception of gender and attending a municipal school. Women have a 11.2% lower chance of being in the top 25%, while studying in a public school increases the chance of being in the bottom 25% by 82%. Attending a private school has a particularly large effect in both directions.

In short, out estimates support our first hypothesis. The clusters influence scores even when accounting for individual, household and school characteristics. Furthermore, estimates also support our second hypothesis. In absolute value, Clusters 5 and 6 (i.e., clusters with high concentrated advantage) produce larger differences than Cluster 1 (a cluster with high concentrated disadvantage). Lastly, we observed that the effect of the spatial context on academic performance is smaller once we account for household, school, and individual characteristics. These results are in line with our third hypothesis and largely demonstrate that the effects of the spatial context are not independent of other spheres.

Additional analyses are presented in the Appendix (Section B). First, we look at our estimates separately by gender. We report larger differences when looking at top and bottom achievement among the high-income cluster. For Cluster 6 – the cluster with the larger effects – the increase in the probability of high achievement is driven solely by men, and the decline in low achievement is much stronger for men. Second, we look at the indirect effects of the clusters on the previous academic achievement, i.e., the effect that the clusters have on test scores in 8th and 4th grade, which in turn affect scores in 10th grade. In general, this analysis provides further evidence in favour of our third hypothesis. Indirect effects are substantive – the gap between Cluster 1 and Cluster 6 goes from 3.6 to 19 points of the SIMCE. As such, indirect effects suggest that the spatial divisions, especially concentrated affluence, are truly important when explaining educational inequalities.

Conclusions

In this study, we examine the effects of individual characteristics, household, school and spatial contexts on 10th grade students' Mathematics SIMCE scores. We focus on Chile, a country that combines high levels of economic inequality with an educational system whose SES-related segregation is among the highest in the world. Specifically, we examined the effects of living in different areas of the Metropolitan Region of Santiago.

We find that the spatial context has an important effect on academic performance. More privileged spatial contexts, especially in terms of wealth concentration, have positive effects on students' Mathematics scores. These effects are equivalent in size to the effects reported for the other contexts considered (household and school) and independent from them. When we consider indirect effects, that is, the impact of the context on students' past performance, we find that the existing educational inequalities across the spatial context increase. This enables us to conclude that space, in general, is a context that exerts an additional (complementary) influence on the processes whereby socioeconomic (dis)advantages are reinforced and accumulated.

This study addresses a critical problem of research on the impact of "area" or "neighbourhood effects" on educational achievement: the lack of an in-depth treatment of the relative nature of the inequality of socioeconomic and spatial opportunities, specifically regarding the potential role of concentrated advantage –compared to the usual emphasis on concentrated disadvantage or poverty. In this regard, we found clear evidence to support the statement that affluence segregation has a larger impact on educational inequality than poverty concentration (Johnson, 2013; Howell, 2019a). More specifically, our findings indicate that intermediate areas display significantly larger differences in academic performance compared to affluent areas than relative to the most impoverished areas of the Metropolitan Region of Santiago. Our results partly support previous research (e.g., Ensminger et al., 1996; Gordon and Monastiriotis, 2006; Howell, 2019b).

Furthermore, we show that the impact of spatial areas is not linear. Average effects are a composition of effects on high achievement and on low achievement. For instance, the more socioeconomically advantaged spatial clusters have a positive impact on academic performance. This impact is explained by both an increase a student's chances of being in the top 25% of SIMCE scores and a decrease their chances of being in the bottom 25%. Conversely, the decrease in average scores from more rural areas marked by poor connectivity and high economic inequality is explained by a strong increase on the likelihood of scoring in the bottom 25%. Along the same line, we show that the beneficial effect of living in wealthy districts is concentrated in men, whereas the disadvantageous effect of residing in vulnerable districts is stronger for women.

Lastly, our findings suggest that spatial mechanisms operate differently when explaining educational inequality across the spatial context. In this study, the more privileged districts display high levels of education and income, along with good accessibility and social networks. Therefore, the positive effect of the spatial context on academic performance may be caused not only by a planned, intentional, and homogeneous parenting style among affluent residents (Lareau, 2011; Lareau and Goyette, 2014), but also by the emergence of types of local cohesion that strengthen forms of socialization focused on the generation of cultural capital. In this regard, social organization and social ties may be operating as tools that the residents of affluent neighbourhoods may be using as an essential part of their privilege (Maloutas and Pantelidou Malouta, 2004; Méndez and Gayo, 2019).

At the other end of the spectrum, we see that the most "disadvantaged" neighbourhoods not only have the highest concentration of residents with low income and schooling levels, but also concentrate experiences of shootings, fights, trash on the streets, infestations, and drugs in their surroundings. Therefore, these spaces may be undergoing a combination of the multiple socialization mechanisms described in the literature (Harding, 2007, 2011) with those associated with those highlighted by environmental theories of neighbourhood (Sharkey and Faber, 2014). This phenomenon may be occurring through the psychological and physical health problems that have been linked to these experiences and their negative impact on cognitive and educational outcomes (Harding, 2009; Caudillo and Torche, 2014; Chen et al., 2018; Shier et al., 2019).

This study has certain strengths and weaknesses worth mentioning and which could prove useful in future research. It should be noted that we considered factors associated not only with the spatial context, but also with the household and school. As we expected, the effects of the spatial context on students' academic performance were considerably reduced when all factors were simultaneously included in the models. This evidence is in line with previous research (Kauppinen, 2008; Owens, 2010; Sykes and Musterd, 2011; Otero et al., 2017) and confirms recent concerns discussed in the literature regarding the need to develop a more integral approach to the multiple contexts in which adolescents participate in order to better determine the specific effect of the spatial context (Nieuwenhuis and Hooimeijer, 2016). In addition, it is relevant to note that we used lagged dependent variables in the regression models, which enabled us to control for past academic performance and capture dynamic effects in the educational achievement process. Thanks to this strategy, we managed to go beyond the associations usually examined in studies on "area" and "neighbourhood effects" (van Ham and Manley, 2012) and address the potential bias of the effects found as a result of non-observed characteristics (Galster and Sharkey, 2017).

Our research opens the door for future research avenues. First, given the nature of our data – particularly the short time frame – we cannot look at the relationship between segregation patterns and the distribution of academic performance. Future research will require long-running longitudinal data to explore changes in both dimensions. Second, future research must look at smaller areas or at different geographical scales to further explore the relevance of the environment. At which geographic scale do we observe larger spatial effects on academic performance? What spatial mechanisms are the main contributing factors to educational inequality at smaller scales? These are relevant challenges for future research that, if tackled, could improve our understanding of spatial divisions and their effects on various outcomes.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459.

Agostini, C. A., Hojman, D., Román, A., & Valenzuela, L. (2016). Segregación Residencial de Ingresos en el Gran Santiago 1992 – 2002: Una Nueva Aproximación Metodológica. *Revista EURE - Revista de Estudios Urbano Regionales*, *42*(127).

Ainsworth, J. W. (2010). Does the Race of Neighborhood Role Models Matter? Collective Socialization Effects on Educational Achievement. *Urban Education*, *45*(4), 401–423.

Andreotti, A., Le Galès, P., & Moreno-Fuentes, F. J. (2015). *Globalised Minds, Roots in the City: Urban Upper-Middle Classes in Europe*. Wiley-Blackwell.

Angotti, T. (Ed.). (2017). *Urban Latin America: Inequalities and Neoliberal Reforms*. Rowman & Littlefield.

Angrist, J., & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ashenfelter, O., & Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics*, *67*(4), 648–660.

Bacqué, M.-H., Bridge, G., Benson, M., Butler, T., Charmes, E., Fijalkow, Y., Jackson, E., Launay, L., & Vermeersch, S. (2015). *The Middle Classes and the City*. Palgrave Macmillan.

Bourdieu, P. (1993). Site Effects. In P. Bourdieu (Ed.), *The Weight of the World: Social Suffering in Contemporary Society* (pp. 123–129). Polity Press.

Bourdieu, P., & Passeron, J.C. (1977). *Reproduction in Education, Society and Culture*. Sage.

Bramley, G., & Karley, N. (2007). Homeownership, Poverty and Educational Achievement: School Effects as Neighbourhood Effects. *Housing Studies*, *22*(5), 693–721.

Bridge, G., Butler, T., & Lees, L. (2012). *Mixed communities: Gentrification by stealth?* Policy Press.

Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Harvard University Press.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do Neighborhoods Influence Child and Adolescent Development? *American Journal of Sociology*, *99*(2), 353–395.

Butler, T., & Hamnett, C. (2007). The Geography of Education: Introduction. *Urban Studies*, *44*(7), 1161–1174. https://doi.org/10.1080/00420980701329174

Caldeira, T. P. (2017). Peripheral urbanization: Autoconstruction, transversal logics, and politics in cities of the global south. *Environment and Planning D: Society and Space, 35*(1), 3–20.

Carrasco, A., & Gunter, H. M. (2019). The "private" in the privatisation of schools: The case of Chile. *Educational Review*, *71*(1), 67–80.

Caudillo, M. L., & Torche, F. (2014). Exposure to Local Homicides and Early Educational Achievement in Mexico. *Sociology of Education*, *87*(2), 89–105.

Chen, S., Guo, C., & Huang, X. (2018). Air Pollution, Student Health, and School Absences: Evidence from China. *Journal of Environmental Economics and Management*, *92*, 465–497.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*, S95–S120.

Contreras, D., Sepúlveda, P., & Bustos, S. (2010). When Schools Are the Ones that Choose: The Effects of Screening in Chile*. *Social Science Quarterly*, *91*(5), 1349–1368.

Crowder, K., & Downey, L. (2010). Interneighborhood Migration, Race, and Environmental Hazards: Modeling Microlevel Processes of Environmental Inequality. *American Journal of Sociology*, *115*(4), 1110–1149.

Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, *1186*, 125–145.

Ensminger, M., Lamkin, R., & Jacobson, N. (1996). School leaving: A longitudinal perspective including neighborhood effects. *Child development*, 67(5), 2400–2416.

Galster, G. C. (2012). The Mechanism(s) of Neighbourhood Effects: Theory, Evidence, and Policy Implications. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Neighbourhood Effects Research: New Perspectives* (pp. 23–56). Springer Netherlands.

Galster, G., & Sharkey, P. (2017). Spatial foundations of inequality: A conceptual model and empirical overview. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *3*(2), 1–33.

Garreton, M. (2017). City profile: Actually existing neoliberalism in Greater Santiago. *Cities*, *65*, 32–50.

Gayo, M., Otero, G., & Méndez, M. L. (2019). Elección escolar y selección de familias: Reproducción de clase media alta en Santiago de Chile. *Revista Internacional de Sociología*, *77*(1), 1–16.

Gingrich, J., & Ansell, B. (2014). Sorting for schools: Housing, education and inequality. *Socio-Economic Review*, *12*(2), 329–351.

Gordon, I., & Monastiriotis, V. (2006). Urban Size, Spatial Segregation and Inequality in Educational Outcomes. *Urban Studies*, *43*(1), 213–236.

Harding, D. J. (2007). Cultural Context, Sexual Behavior, and Romantic Relationships in Disadvantaged Neighborhoods. *American Sociological Review*, *72*(3), 341–364.

Harding, D. J. (2009). Collateral Consequences of Violence in Disadvantaged Neighborhoods. *Social Forces; a Scientific Medium of Social Study and Interpretation*, *88*(2), 757–784.

Harding, D. J. (2011). Rethinking the Cultural Context of Schooling Decisions in Disadvantaged Neighborhoods From Deviant Subculture to Cultural Heterogeneity. *Sociology of Education*, *84*(4), 322–339.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer. Available at: http://web.stanford.edu/~hastie/ElemStatLearn/

Howell, J. (2019a). The unstudied reference neighborhood: Towards a critical theory of empirical neighborhood studies. *Sociology Compass*, *13*(1), e12649.

Howell, J. (2019b). The Truly Advantaged: Examining the Effects of Privileged Places on Educational Attainment. *The Sociological Quarterly*, *60*(3), 420–438.

Husson, F., Lê, S., & Pagès, J. (2011). *Exploratory multivariate analysis by example using R*. Chapman & Hall/CRC.

Jencks, C., & Mayer, S. (1990). The Social Consequences of Growing up in a Poor Neighborhood. In L. Lynn & M. McGeary (Eds.), *Inner City Poverty in the US* (pp. 111–185). National Academy Press.

Johnson, O. (2013). Is Concentrated Advantage the Cause? The Relative Contributions of Neighborhood Advantage and Disadvantage to Educational Inequality. *The Urban Review*, *45*(5), 561–585.

Kauppinen, T. M. (2008). Schools as Mediators of Neighbourhood Effects on Choice Between Vocational and Academic Tracks of Secondary Education in Helsinki. *European Sociological Review*, *24*(3), 379–391.

Lamont, M., Beljean, S., & Clair, M. (2014). What is missing? Cultural processes and causal pathways to inequality. *Socio-Economic Review*, *12*(3), 573–608.

Lareau, A. (2011). *Unequal childhoods: Class, race, and family life*. University of California Press.

Lareau, A., & Goyette, K. (Eds.). (2014). *Choosing homes, choosing schools*. Russell Sage Foundation.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, *126*(2), 309–337.

Maloutas, T., & Pantelidou Malouta, M. (2004). The glass menagerie of urban governance and social cohesion: Concepts and stakes/concepts as stakes. *International Journal of Urban and Regional Research*, *28*(2), 449–465.

Marcuse, P. (2005). Enclaves yes, ghettos no. In D. Varady (Ed.), *Desegregating the city: Ghettos, enclaves, and inequality* (pp. 15–30). State University of New York Press.

Massey, D. S. (1996). The age of extremes: Concentrated affluence and poverty in the twenty-first century. *Demography*, *33*(4), 395–412.

Méndez, M. L., Otero, G., Link, F., López-Morales, E., & Gayo, M. (2020). Neighbourhood cohesion as a form of privilege. *Urban Studies* (Online first).

Méndez, M.-L., & Gayo, M. (2019). *Upper Middle Class Social Reproduction: Wealth, Schooling, and Residential Choice in Chile*. Palgrave Pivot Series.

Mizala, A., & Torche, F. (2012). Bringing the schools back in: The stratification of educational achievement in the Chilean voucher system. *International Journal of Educational Development*, *32*(1), 132–144.

Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence. *Criminology*, *39*(3), 517–558.

Musterd, S., Andersson, R., & Galster, G. (2019). Neighbourhood effects on social outcomes. In T. Schwanen & R. Van Kempen (Eds.), *Handbook of Urban Geography* (pp. 281–296). Elgar.

Musterd, S., Marcińczak, S., Ham, M. van, & Tammaru, T. (2017). Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. *Urban Geography*, 38(7), 1062–1083.

Nieuwenhuis, J., & Hooimeijer, P. (2016). The association between neighbourhoods and educational achievement, a systematic review and meta-analysis. *Journal of Housing and the Built Environment*, *31*(2), 321–347.

Núñez, J., & Tartakowsky, A. (2011). The relationship between income inequality and inequality of opportunities in a high-inequality country: The case of Chile. *Applied Economics Letters*, *18*(4), 359–369.

OECD (2019). Income inequality (indicator). doi: 10.1787/459aa7f1-en (Accessed on 22 December 2019).

Otero, G., Carranza, R., & Contreras, D. (2017). 'Neighbourhood effects' on children's educational achievement in Chile: The effects of inequality and polarization. *Environment and Planning A, 49*(11), 2595–2618.

Owens, A. (2010). Neighborhoods and Schools as Competing and Reinforcing Contexts for Educational Attainment. *Sociology of Education*, *83*(4), 287–311.

Pinçon-Charlot, M., & Pinçon, M. (2018). Social Power and Power Over Space: How the Bourgeoisie Reproduces itself in the City. *International Journal of Urban and Regional Research*, *42*(1), 115–125.

Rose, H., & Betts, J. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, 86(2), 497–513.

Ruiz-Tagle, J., & López-Morales, E. (2014). El estudio de la segregación residencial en Santiago de Chile: Revisión crítica de algunos problemas metodológicos y conceptuales. *EURE (Santiago), 40*(119), 25–48.

Sampson, R. (2019). Neighbourhood effects and beyond: Explaining the paradoxes of inequality in the changing American metropolis. *Urban Studies, 56*(1), 3-32.

Sampson, R., & Groves, W. B. (1989). Community Structure and Crime: Testing Social-Disorganization Theory. *American Journal of Sociology*, *94*(4), 774–802.

Sampson, R., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science*, *277*(5328), 918–924.

Santos, H., & Elacqua, G. (2016). Socioeconomic school segregation in Chile: Parental choice and a theoretical counterfactual analysis. *CEPAL Review*, *119*, 123–137.

Savage, M., Bagnall, G., & Longhurst, B. (2005). *Globalisation and belonging*. Sage.

Sharkey, P. (2010). The acute effect of local homicides on children's cognitive performance. *Proceedings of the National Academy of Sciences*, *107*(26), 11733–11738.

Sharkey, P., & Faber, J. W. (2014). Where, When, Why, and For Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects. *Annual Review of Sociology*, *40*(1), 559–579.

Shier, V., Nicosia, N., Shih, R., & Datar, A. (2019). Ambient air pollution and children's cognitive outcomes. *Population and Environment*, *40*(3), 347–367.

Slater, T. (2013). Your Life Chances Affect Where You Live: A Critique of the 'Cottage Industry' of Neighbourhood Effects Research. *International Journal of Urban and Regional Research*, *37*(2), 367–387.

Small, M. L., & Newman, K. (2001). Urban Poverty after The Truly Disadvantaged: The Rediscovery of the Family, the Neighborhood, and Culture. *Annual Review of Sociology*, *27*(1), 23–45.

Sykes, B., & Musterd, S. (2011). Examining Neighbourhood and School Effects Simultaneously: What Does the Dutch Evidence Show? *Urban Studies*, *48*(7), 1307–1331.

Valenzuela, J. P., Bellei, C., & de los Ríos, D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, *29*(2), 217–241.

van Ham, M., Manley, D., Bailey, N., Simpson, L., & Maclennan, D. (2012). Neighbourhood Effects Research: New Perspectives. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Neighbourhood Effects Research: New Perspectives* (pp. 1–21). Springer Netherlands.

Wacquant, L. (2008). *Urban outcasts: A Comparative Sociology of Advanced Marginality*. Policy Press.

Wilson, W. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

Woolley, M., & Grogan-Kaylor, A. (2006). Protective Family Factors in the Context of Neighborhood: Promoting Positive School Outcomes. *Family Relations*, *55*(1), 93–104.

Zimmerman, S. D. (2019). Elite Colleges and Upward Mobility to Top Jobs and Top Incomes. *American Economic Review*, 109(1), 1–47.

Appendix

SECTION A: Measuring spatial context: Hierarchical Clustering on Principal Components.

In order to identify the spatial clusters, we use Hierarchical Clustering on Principal Components (HCPC), which is particularly useful when the number of variables is large and their correlation is high, as in this case. This approach uses a three-step algorithm that (1) reduces the number of dimensions (making the analysis less susceptible to particular variables), (2) groups by similarity, and (3) splits groups by dissimilarity (Husson et al., 2011). First, it uses principal component analysis (PCA) to reduce the number of dimensions from the number of variables to two (or more) orthogonal components that reflect different aspects of these variables. The next step is to use hierarchical clustering on the principal components. This algorithm groups each district into clusters by ordering their principal components by similarity, iterating the process until all districts are part of a single cluster.[5] The last step uses K-means clustering to improve the previous partition. Concretely, K-means clustering takes the partition given in step 2 as its starting point and splits clusters that show high dissimilarity.[6]

Our HCPC analysis considers 28 district-level variables, including the average years of schooling of household heads of the district, poverty rate, monthly household income, unemployment rate, economic inequality based on the GINI index, access to education, air pollution, experiences of violence, territorial stigmatization, and socioeconomic polarization. Table A1 shows the whole set of variables used in the analysis and its description.

<<TABLE A1 ABOUT HERE>>

The HCPC approach allowed us to identify six clear clusters of districts. Table A2 describes the variables per cluster. As shown in Table A2, Cluster 1 includes most low-income districts. It includes 13 districts and 17.5% of all students in the sample. It is characterized by low inequality (suggesting a homogeneous cluster), and a low educational level (considering both years of schooling and the share of residents with higher education). It is also characterized by experiences of violence: a higher rate of fights, shootings, and drug consumption, as well as trash, graffiti, and infestations. Although these districts are within the urban part of the Region, they are mostly located in the periphery (Garretón,

[5] A distance metric is required to define similarity. HCPC uses Ward's criterion in the Hierarchical clustering step, as this variance minimizing method is also used in the PCA step.
[6] See chapters 2 and 3 from Husson et al. (2011) for a more detailed description of the HCPC algorithm.

2017). Table A2 shows that this cluster is also characterised by the lowest average income, the highest share of people reporting territorial stigma, and together with Cluster 3, the highest poverty rate. In brief, it clearly represents the condition of living in marginalised areas.

Clusters 2 and 3 identify the rural districts located in the periphery of the Metropolitan Region of Santiago. Cluster 2 is the smallest one (4 districts and 0.5% of all students). They both share low levels of access to services: cash machines, supermarkets, and community services in the case of Cluster 2 and parks, education, and pharmacies for Cluster 3. They also show high relative levels of polarization, which suggests the existence of socioeconomic inequalities within them. Roughly speaking, the periphery districts –as many rural districts in Chile and Latin America– are characterized by two large and relatively isolated groups of people: high-income landowners and low-income unskilled workers (Ruiz-Tagle and López, 2014). The remaining characteristic variables differ between the two clusters. Cluster 2 is also characterized by a higher share of people reporting infestations (e.g. rats, mice, fleas) and larger networks, which reinforces the idea of a rural and polarized cluster. Cluster 3 is characterized by relatively high participation in neighbourhood associations and higher poverty rates.

Cluster 4 comprises middle-class districts. It is by far the largest cluster in terms of students (38.8%) and includes 14 of the 52 districts. Being the most populous, it includes several urban centres and is characterized by access to services such as transport, education, and supermarkets. It also shows a low rate of people perceiving a stigma associated with living in those districts. It shows a relatively low level of inequality, measured through the Gini index, suggesting a relatively homogeneous cluster (confirmed by the low standard deviation for income). It also shows a relatively lower share of people with higher education and lower air contamination. Being the most populous cluster, its averages are similar to the sample average, which explains why the standard deviations presented in Table A2 are relatively small.

Cluster 5 represents the upper middle-class. It includes 5 districts and 18.5% of all students. It is characterized by relatively low poverty, stigma, and participation in neighbourhood organizations. On the other hand, it is a much more heterogeneous district, as shown by a high Gini index and a high standard deviation for income. This district includes the largest urban centres in the region, which could explain the higher levels of acoustic and visual pollution.

Cluster 6 captures the higher income districts of the Metropolitan Region of Santiago and the whole of Chile (Méndez and Gayo, 2019). It is comprised of 4 districts and is the second smallest cluster, with 10% of all students. It is characterized by high average income (with a positive deviation of 2.6 standard deviations relative to the overall mean) and high education, both in years of schooling (12.8 versus 10.2) and the share of people with tertiary education (35% versus 24%). It is also characterized by low levels of infestations, trash lying around, and fights in the street. The cluster also shows a low poverty rate (less than 1%).

To gain a clearer understanding of the areas identified, Figure A2 shows the geographical location of the six spatial clusters in the Metropolitan Region of Santiago. As can be observed, the poorer districts belonging to Cluster 1 are mostly located in the southern/western outskirts of Santiago. The districts located in clusters 2 and 3, characterised by a lack of connectivity, are located in the more rural peripheral areas of the Metropolitan Region of Santiago. The middle class districts that define Cluster 4 are more variedly distributed across the urban geography of the region, while those that constitute Cluster 5 are mostly located in the city centre. The 4 districts that represent Cluster 6 are clearly located in the northeastern part of the city. This area of the city is colloquially known as the *cono de alta renta* (high-income cone).

<<TABLE A2 ABOUT HERE>>

SECTION B: Complementary analyses

Disaggregated models by gender

Table A3 shows the models disaggregated by gender. There is a relatively homogeneous effect for both men and women. When we look at the average test score (models 1 and 2), we see that the effect of being in each cluster has the same sign for men and women except for Cluster 1, where the positive effect is solely driven by women. Quantitatively, women show a stronger effect in clusters 4 and 5, while men have a stronger effect in cluster 6.

<<TABLE A3 ABOUT HERE>>

The main gender differences can be observed when analysing high and low achievement. Cluster 2 shows a negative effect for women and a positive effect for men when predicting high achievement (i.e., being in the top 25% of the SIMCE distribution in Models 3 and 4). On the other hand, Clusters 4 and 5 show similar effects between men and women, both significant and positive. The positive effect of being in Cluster 6 is driven solely by men, who have a 22.3% higher chance of being in the top 25%. In other words, the positive effect of concentrated advantage holds only for men.

With respect to low achievement, the negative effect of Clusters 4 to 6 holds for both men and women. Cluster 2 also shows small gender differences, but a positive effect. Given the bimodal nature of Cluster 2, we can conclude that the beneficial effect (i.e., increasing the chance of being in the top 25%) is driven solely by men, while the detrimental effect (i.e., increasing the chance of being in the bottom 25%) is common across genders. Lastly, Cluster 1 reduces the chances of being in the bottom 25% for women but has no effect for men. Overall, we see a picture of increasing gender inequalities for clusters with a concentration of high-income households, namely Clusters 2 and 6, as men in these clusters have a higher chance of scoring higher than women.

The indirect effect of household, school, and spatial contexts

It could be argued that our estimates are in fact lower bound estimates of these whole effects, as they do not take into account these indirect effects. Figure A1 illustrates the expected causal channels between test scores and the rest of the control variables. In this section, we aim to block the indirect effect (b), thus ‘shifting’ it so that it is included in (a). In that way, the whole effect of our independent variables will be included.

<<FIGURE A1 ABOUT HERE>>

In order to do this, we follow a three-steep procedure. We first estimate the effect of household, school, and neighbourhood factors on our earliest test score (4th grade) and then take the predicted residual $\hat{\varepsilon}_{ijk,t-2}$, as shown in equation 3. The following step is to estimate the effect of the next test score (8th grade), controlling for the predicted residual for the 4th grade score and the other factors, as shown in equation 4, proceeding to take its predicted residual $\hat{\varepsilon}_{ijk,t-1}$. Lastly, we estimate a simile of equation 1 with the predicted residuals instead of the actual past test scores, as shown in equation 5. This equation accounts for the complete effect on 10th grade scores, as previous test scores are 'clean' of household, school, and neighbourhood effects.

$$T_{ijk,t-2} = \alpha_1 + \beta_1 X_{it} + \gamma_1 S_{jt} + C_{1_{kt}} + \varepsilon_{ijk,t-2}. \tag{3}$$

$$T_{ijk,t-1} = \alpha_2 + \beta_2 X_{it} + \gamma_2 S_{jt} + C_{2_{kt}} + \eta_2 \hat{\varepsilon}_{ijk,t-2} + \varepsilon_{ijk,t-1}. \tag{4}$$

$$T_{ijk,t} = \alpha_3 + \beta_3 X_{it} + \gamma_3 S_{jt} + C_{3_{kt}} + \delta_3 \hat{\varepsilon}_{ijk,t-2} + \eta_3 \hat{\varepsilon}_{ijk,t-1} + \varepsilon_{ijkt}. \tag{5}$$

Table A4 presents these results. Models 1 and 2 are equations 3 and 4 respectively. These equations look at the effect of all factors on the previous SIMCE test scores (4th and 8th grade). Equations 3 and 4 compare the two approaches to estimate the effect of each cluster. Model 3 is the standard regression (Model 4 in table 4 and equation 1), while Model 4 is equation 5, the same model but using the residual SIMCE scores obtained from Models 1 and 2 in Table A4.

<<TABLE A4 ABOUT HERE>>

Model 1 shows that, on average, only Clusters 5 and 6 display a higher 4th-grade SIMCE score. A similar effect can be seen for school type, as only private schools show a significant and higher score on average. Model 1 shows that school achievement starts to branch out between higher status students and the rest even at a young age. Model 2 shows similar results for 8th-grade SIMCE, with the addition that effects become larger –4 years after the first test, individual-, school-, and neighbourhood-level factors become more important when explaining score gaps.

The main results stem from comparing models 3 and 4. Model 3 is the benchmark model, while Model 4 uses residual test scores. The difference between the two is the extent to which the factors considered indirectly affect 10th-grade SIMCE scores (i.e., the effect of individual, household, school, and neighbourhood factors on 4th and 8th grade scores, which in turn affect 10th grade SIMCE scores). First, we see that almost all coefficients increase, which means that the indirect effect has the same direction as the main effect. The only exception is Cluster 1, which shows no difference between models. This increase is monotonic with respect to the clusters –the higher the socioeconomic level of the cluster, the larger the increase in the coefficient. The coefficient of Cluster 2 increases 1.7 times, while for Cluster 6 we see a 4.3-fold increase. The overall gap between areas representing concentrated advantage and disadvantage (Cluster 6 relative to Cluster 1) increases from 3.6 SIMCE points to over 19 points. The gaps in academic performance between the most affluent spatial cluster and the intermediate areas are also noticeably increased. These findings provide clearer support for our second hypothesis. Indirect effects increase existing inequalities in the spatial context, especially those referring to concentrated advantage.

Interestingly, the effect of the neighbourhood clusters also increases relative to individual and school level factors. The effect of studying in Cluster 6 schools in Model 3 is equivalent to around 5.5 years of additional education of one of the parents or a 2.6% increase in household income. Regarding the indirect effect of attending Cluster 6 schools, it is equivalent to around 7.4 years of additional education of one of the parents or a 3.4% increase in household income. When this indirect effect is taken into account, individual, school, and neighbourhood variables increase their size. However, it is neighbourhood variables that show the largest increase of all. These results provide further evidence in favour of our third hypothesis (the effects of the spatial context on academic performance are reduced when household and school contexts are included). However, our findings also indicate that the independent effects of the spatial context are larger than initially reported.

Tables

Table 1. Number of observations

	Number of cases	Share of total	Share of last step	Average SIMCE
Total	248,158	100%		-
SIMCE Students' test (10th grade)	196,344	79.10%	79.10%	265.51
SIMCE Parents' questionnaire	162,779	65.60%	82.90%	267.72
Household information	141,934	57.20%	87.20%	269.01
Duplicates	115,970	46.70%	81.70%	275.12
SIMCE Students' test (8th grade)	115,913	46.70%	100%	275.12
SIMCE Students' test (4th grade)	101,792	41.00%	87.80%	279.38
CASEN 2015	101,655	41.00%	99.90%	279.42
The Metropolitan Region of Santiago	37,000	14.90%	36.40%	286.12

Table 2. Summary of spatial clusters.

	Description
Cluster 1	Low-income districts, with low economic inequality, and a low educational level. It is also characterized by experiences of violence (e.g., fights, shootings, and drug consumption). This cluster comprises 13 districts including El Bosque, La Pintana, Lo Espejo, and Pudahuel.
Cluster 2	Rural districts characterised by low levels of access to services, especially cash machines, supermarkets, and community services. High levels of economic polarization. It is also characterised by a higher share of people reporting infestations. and larger networks. This cluster comprises 4 districts: Alhue, Calera de Tango, Maria Pinto, and San Pedro.
Cluster 3	Rural districts characterised by low levels of access to services, especially parks, educational institutions, and pharmacies. High levels of economic polarization. It is also characterised by relatively high participation in neighbourhood associations and higher poverty rates. This cluster comprises 12 districts, including Buin, Colina, Pirque, and Tiltil.
Cluster 4	Middle-class districts characterised by high accessibility to services such as transport, education, and supermarkets. A relatively low level of inequality. It shows a lower share of people with higher education, and relatively low levels of perceived residential stigma. This cluster comprises 14 districts, including Independencia, La Florida, Puente Alto, Quinta Normal, and Renca.
Cluster 5	Upper middle-class districts characterised by low levels of poverty, and residential stigma. Relatively high levels of economic inequality. Participation in neighbourhood organizations. Acoustic and visual pollution. This cluster comprises 5 districts: La Reina, Macul, Ñuñoa, San Miguel, and Santiago.
Cluster 6	This cluster captures the higher income districts of the Metropolitan Region of Santiago. It is also characterized by high levels of education, both in years of schooling and the share of people with tertiary education. Low levels of infestations, trash lying around, and fights in the street. It also shows a low poverty rate (less than 1%). This cluster comprises 4 districts: Las Condes, Lo Barnechea, Providencia, and Vitacura.

Table 3. Descriptive statistics for individual, household, and school variables.

	Min	Max	Mean	SD
Students' characteristics				
SIMCE 4th grade	106.0	356.5	284.6	45.9
SIMCE 8th grade	145.2	402.7	277.9	46.7
SIMCE 10th grade	97.3	425.7	286.1	63.8
Female	0	1	0.52	
Household context				
Father's years of schooling	0	22	12.70	3.80
Mother's years of schooling	0	22	12.70	3.50
Monthly household income (log)	10.82	14.56	12.97	0.71
School context (type of school)				
Public school	0	1	0.18	
Subsidised school	0	1	0.68	
Private school	0	1	0.14	

Table 4. Regression models of academic achievement in mathematics.

	SIMCE test score in 10th grade (OLS)				Logit Top 25	Logit Bottom 25
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Spatial context						
Cluster 1	-1.348***	1.341***	1.682***	0.949	1.027	0.859***
	(0.000)	(0.042)	(0.146)	(0.563)	(0.028)	(0.019)
Cluster 2	-3.397***	-6.196***	-7.577***	-7.265***	1.214***	1.497***
	(0.000)	(0.059)	(0.345)	(0.994)	(0.086)	(0.076)
Cluster 3	(Ref)	(Ref)	(Ref)	(Ref)	(Ref)	(Ref)
Cluster 4	11.767***	4.581***	3.086***	2.590***	1.142***	0.846***
	(0.000)	(0.122)	(0.226)	(0.386)	(0.019)	(0.013)
Cluster 5	20.596***	5.247***	2.643***	2.806***	1.110***	0.821***
	(0.000)	(0.216)	(0.436)	(0.381)	(0.029)	(0.008)
Cluster 6	58.844***	14.288***	7.261***	4.598**	1.177**	0.651***
	(0.000)	(0.656)	(1.418)	(1.420)	(0.086)	(0.032)
Control variables						
Household context						
Father's years of schooling			1.090***	0.830**	1.042***	0.971***
			(0.247)	(0.230)	(0.009)	(0.009)
Mother's years of schooling			1.106***	0.846***	1.035**	0.963***
			(0.189)	(0.149)	(0.015)	(0.003)
Monthly household income (log)			1.137	1.785**	1.052	0.892***
			(0.908)	(0.598)	(0.060)	(0.025)
School context (type of school)						
Municipal-public schools				-7.941*	0.976	1.820***
				(3.109)	(0.069)	(0.228)
Subsidised-private schools				(Ref)	(Ref)	(Ref)
Private schools				8.448***	1.491***	0.489***
				(1.870)	(0.163)	(0.052)
Students' characteristics						
Female			-0.740	-1.065	0.888**	1.027
			(0.757)	(0.741)	(0.048)	(0.033)
SIMCE 4th grade		0.167***	0.152***	0.153***	1.006***	0.994***
		(0.014)	(0.014)	(0.016)	(0.001)	(0.000)
SIMCE 8th grade		0.920***	0.875***	0.866***	1.047***	0.964***
		(0.019)	(0.023)	(0.024)	(0.001)	(0.001)
Constant	272.144***	-21.559***	-45.220**	-43.859***	0.000***	269,987.305***
	(0.000)	(4.089)	(11.744)	(6.890)	(0.000)	(65,185.823)
Number of students	37,000	37,000	37,000	37,000	37,000	37,000
R-squared	0.069	0.582	0.593	0.597		

Notes: Clustered standard errors in parentheses. Logits are odd-ratios.

*** p<0.01, ** p<0.05, * p<0.1

Figures

Figure 1. Factor map representing the two principal components and the spatial context (clustered districts).

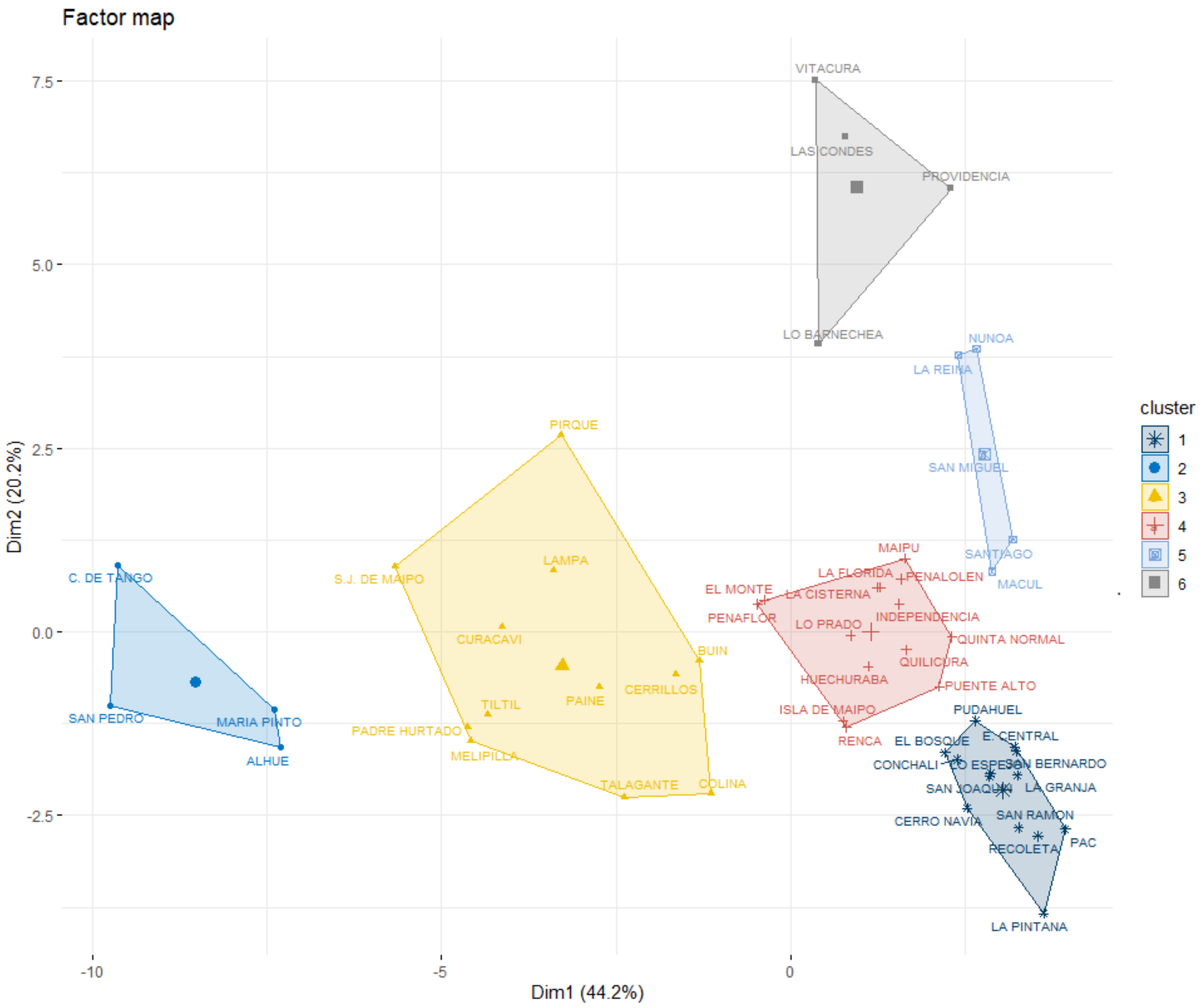

Figure 2. Kernel density of tests scores by cluster.

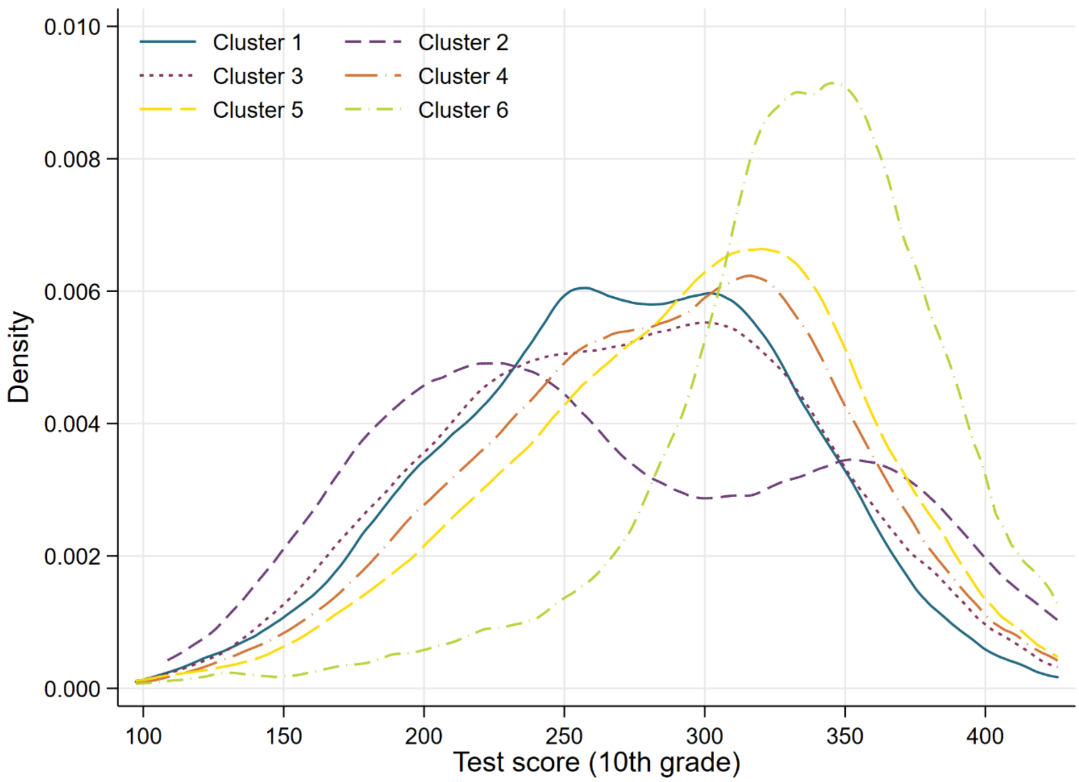

Figure 3. Relationship between previous test score and high achievement, by cluster.

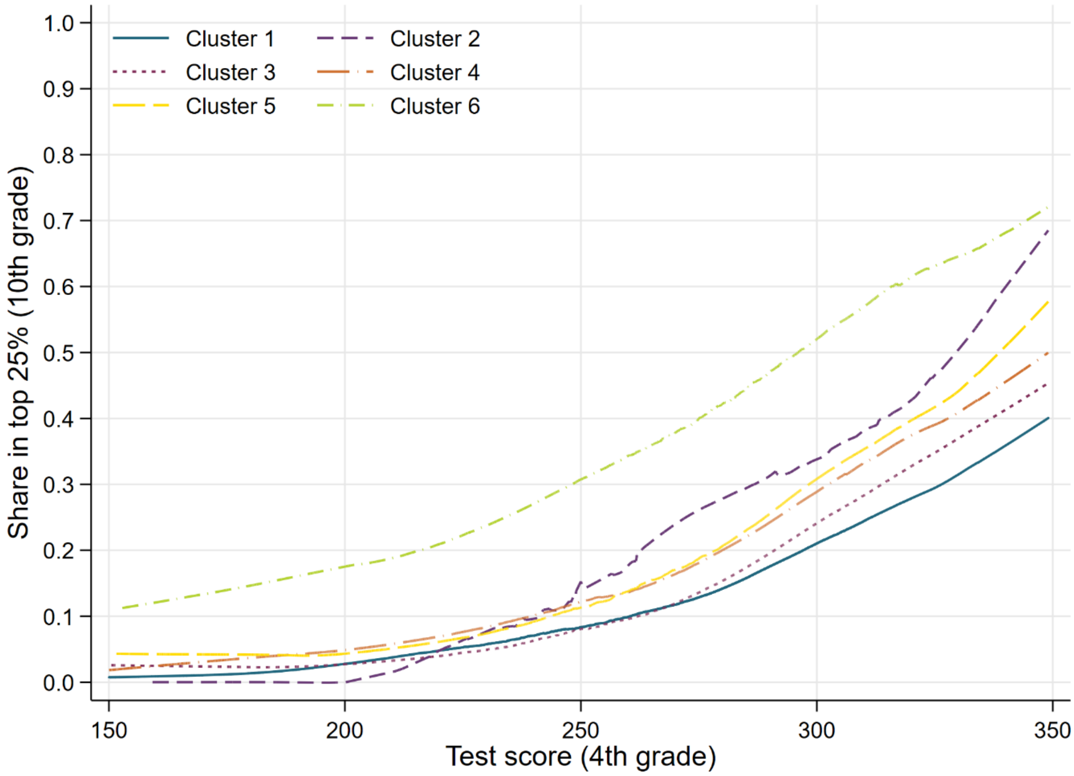

Table A1. Descriptive statistics of the spatial variables.

	Min	Max	Mean	SD	Description
Income (CLP)	321,269	563,024	430,534	63,771	Total household income (per capita)
Schooling	10.22	12.70	11.87	0.53	Years of schooling (0-22)
Higher ed. (share)	23.9%	36.5%	32.1%	0.03	Share with complete higher degree
Unemployment	2.4%	7.2%	5.0%	0.01	Unemployment rate
Poverty	3.0%	10.6%	6.8%	0.02	Absolute poverty rate
Gini	0.31	0.46	0.38	0.04	Income inequality (Gini index)
Polarization	0.0001	0.0017	0.0003	0.0003	Income polarization (Esteban & Ray, 1994)
Networks	9.88	10.63	10.19	0.20	Sum of 'yes' answers to "Do you know someone that could help in (11 different tasks)"
Participates in neighbourhood	1.3%	6.5%	3.3%	0.01	Participates in neighbourhood organizations
Experiences fights	34.1%	63.3%	50.4%	0.08	Experiences people fighting in the street (in a 15 minute walk radius)
Experiences shootings	28.0%	62.8%	42.6%	0.10	Experiences shootings in the street (in a 15 minute walk radius)
Trash in surroundings	30.6%	62.8%	47.5%	0.08	Has seen accumulated trash in the street (in a 15 minute walk radius)
Graffiti in surroundings	37.1%	70.9%	53.1%	0.08	Has seen graffiti on cars or houses (in a 15 minute walk radius)
Plagues in surroundings	32.5%	65.4%	47.1%	0.09	Has seen plagues of insects, rats, dogs, bats, etc. (in a 15 minute walk radius)
Drugs in surroundings	56.1%	80.7%	70.8%	0.07	Experiences drug trafficking in the street (in a 15 minute walk radius)
Stigma for residence	0.2%	1.8%	0.8%	0.00	Has been discriminated or treated unfairly due to the place they live
Access to transport	94.5%	99.8%	97.9%	0.01	Lives less than 8 blocks away from public transport
Access to education	93.9%	99.5%	96.8%	0.02	Lives less than 20 blocks away from a school or preschool
Access to health	82.8%	97.4%	89.8%	0.05	Lives less than 20 blocks away from a health centre
Access to supermarket	90.7%	99.3%	96.8%	0.02	Lives less than 20 blocks away from a supermarket or store
Access to cash machines	85.6%	98.1%	93.1%	0.04	Lives less than 20 blocks away from a cash machine (ATM)
Access to sports	86.7%	97.6%	92.9%	0.03	Lives less than 20 blocks away from a sport centre
Access to parks	87.0%	98.7%	94.8%	0.03	Lives less than 20 blocks away from a park
Access to community	85.0%	98.1%	91.2%	0.03	Lives less than 20 blocks away from a community centre
Access to pharmacy	66.8%	95.7%	86.7%	0.05	Lives less than 20 blocks away from a drugstore or pharmacy
Acoustic pollution	37.8%	65.5%	55.6%	0.08	Experiences noise pollution (in a 15 minute walk radius)
Air pollution	32.7%	61.6%	47.6%	0.09	Experiences air pollution (in a 15 minute walk radius)
Visual pollution	22.2%	52.8%	36.6%	0.09	Experiences visual pollution (in a 15 minute walk radius)

Table A2. Summary of district-level variables by spatial cluster (means).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Total
Number of students	6490	191	5462	14339	6827	3691	37000
Number of districts	13	4	12	14	5	4	52
Income (CLP)	196,845	186,313	211,070	215,394	428,033	923,943	321,269
Schooling	9.61	8.40	9.35	9.15	12.41	12.75	10.22
Higher ed. (share)	21.7%	22.4%	22.9%	19.4%	30.7%	34.7%	23.9%
Unemployment	2.8%	0.4%	2.5%	2.0%	1.9%	4.1%	2.4%
Poverty	6.2%	0.9%	2.3%	3.1%	2.2%	0.0%	3.0%
Gini	28.7%	23.9%	27.6%	26.7%	42.1%	35.9%	30.9%
Polarization	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Networks	9.61	9.88	9.74	9.90	9.87	10.54	9.88
Participates in neighbourhood	0.6%	2.8%	2.3%	0.9%	1.3%	2.8%	1.3%
Experiences fights	63.2%	5.8%	20.4%	32.6%	37.1%	4.7%	34.1%
Experiences shootings	49.9%	0.0%	1.3%	37.2%	22.7%	4.2%	28.0%
Trash in surroundings	53.8%	5.2%	9.1%	26.6%	46.3%	9.3%	30.6%
Graffiti in surroundings	45.6%	2.2%	11.7%	35.8%	49.9%	42.9%	37.1%
Plagues in surroundings	49.5%	11.2%	26.8%	29.5%	39.2%	11.6%	32.5%
Drugs in surroundings	77.4%	22.6%	20.9%	59.1%	67.8%	39.4%	56.1%
Stigma for residence	0.3%	0.0%	0.0%	0.0%	0.3%	0.9%	0.2%
Access to transport	98.2%	68.5%	84.6%	93.8%	98.6%	99.3%	94.5%
Access to education	98.5%	64.2%	75.5%	96.1%	97.5%	99.0%	93.9%
Access to health	93.4%	52.0%	59.7%	84.0%	90.6%	81.1%	82.8%
Access to supermarket	95.5%	61.5%	72.2%	90.2%	98.7%	98.3%	90.7%
Access to cash machines	92.4%	42.2%	62.1%	83.9%	96.4%	97.9%	85.6%
Access to sports	88.2%	55.1%	68.1%	89.4%	94.0%	89.9%	86.7%
Access to parks	91.4%	51.6%	62.1%	88.1%	94.8%	99.5%	87.0%
Access to community	91.0%	30.5%	67.4%	84.0%	95.0%	88.4%	85.0%
Access to pharmacy	84.6%	2.6%	29.8%	55.3%	87.9%	98.7%	66.8%
Acoustic pollution	38.4%	12.3%	24.1%	34.8%	51.6%	44.6%	37.8%
Air pollution	35.5%	15.8%	24.2%	30.2%	48.3%	22.2%	32.7%
Visual pollution	27.0%	3.1%	6.5%	17.4%	43.3%	17.3%	22.2%

Note: Results are weighted by the number of students in the regression analysis.

Table A3. Disaggregated regression models of academic achievement in mathematics by gender.

	OLS test score		Logit top 25		Logit bottom 25	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Spatial context						
Cluster 1	2.072***	-0.283	1.023	1.028	0.783***	0.958
	(0.510)	(0.686)	(0.035)	(0.025)	(0.012)	(0.030)
Cluster 2	-7.063***	-7.706***	0.961	1.486***	1.896***	1.289***
	(1.060)	(0.978)	(0.078)	(0.092)	(0.075)	(0.076)
Cluster 3	(Ref)	(Ref)	(Ref)	(Ref)	(Ref)	(Ref)
Cluster 4	3.892***	1.180**	1.206***	1.083***	0.824***	0.875***
	(0.390)	(0.417)	(0.025)	(0.018)	(0.007)	(0.023)
Cluster 5	3.990***	1.475**	1.070***	1.162***	0.763***	0.887***
	(0.358)	(0.570)	(0.027)	(0.036)	(0.008)	(0.009)
Cluster 6	4.403**	4.938**	1.113	1.239***	0.712***	0.588***
	(1.391)	(1.477)	(0.093)	(0.085)	(0.038)	(0.028)
Control variables						
Household context						
Father's years of schooling	0.752**	0.916**	1.036***	1.046***	0.973***	0.969***
	(0.206)	(0.292)	(0.009)	(0.010)	(0.010)	(0.010)
Mother's years of schooling	0.909***	0.776***	1.036*	1.033**	0.960***	0.968***
	(0.186)	(0.117)	(0.019)	(0.014)	(0.007)	(0.003)
Monthly household income (log)	2.040***	1.515	1.052	1.048	0.886**	0.900***
	(0.427)	(0.992)	(0.059)	(0.069)	(0.054)	(0.028)
School context (type of school)						
Municipal-public schools	-7.836**	-7.993*	1.105	0.868*	1.815***	1.841***
	(2.996)	(3.438)	(0.108)	(0.064)	(0.203)	(0.274)
Subsidised-private schools	(Ref)	(Ref)	(Ref)	(Ref)	(Ref)	(Ref)
Private schools	8.899***	7.925**	1.477***	1.512***	0.383***	0.627***
	(1.700)	(2.411)	(0.148)	(0.223)	(0.050)	(0.063)
Students' characteristics						
SIMCE 4th grade	0.159***	0.146***	1.007***	1.006***	0.993***	0.994***
	(0.012)	(0.020)	(0.001)	(0.001)	(0.000)	(0.001)
SIMCE 8th grade	0.855***	0.880***	1.047***	1.047***	0.963***	0.966***
	(0.027)	(0.022)	(0.001)	(0.001)	(0.002)	(0.001)
Constant	-47.786***	-41.262***	0.000***	0.000***	480,107.890***	147,533.233***
	(5.513)	(8.847)	(0.000)	(0.000)	(195,888.694)	(37,180.393)
Number of students	19,337	17,663	19,337	17,663	19,337	17,663
R-squared	0.611	0.58				

Notes: Clustered standard errors in parentheses. Logits are odd-ratios.

*** p<0.01, ** p<0.05, * p<0.1

Table A4. Regressions accounting for the indirect effect of all factors on the clusters

	4th grade	8th grade	10th grade	10th grade
	Model 1	**Model 2**	**Model 3**	**Model 4**
Spatial context				
Cluster 1	-2.042	0.304	0.949	0.901
	(1.030)	(0.370)	(0.563)	(0.543)
Cluster 2	-4.006	-5.458***	-7.265***	-12.605***
	(2.014)	(0.583)	(0.994)	(0.978)
Cluster 3	(Ref)	(Ref)	(Ref)	(Ref)
Cluster 4	-0.087	4.582***	2.590***	6.546***
	(0.795)	(0.309)	(0.386)	(0.338)
Cluster 5	6.331***	7.088***	2.806***	9.912***
	(0.223)	(0.193)	(0.381)	(0.369)
Cluster 6	10.130***	15.895***	4.598**	19.915***
	(1.329)	(0.851)	(1.420)	(1.412)
Control variables				
Household context				
Father's years of schooling	1.476***	1.888***	0.830**	2.691***
	(0.064)	(0.086)	(0.230)	(0.227)
Mother's years of schooling	1.497***	1.911***	0.846***	2.730***
	(0.112)	(0.090)	(0.149)	(0.132)
Monthly household income (log)	4.011***	4.059***	1.785**	5.914***
	(0.525)	(0.673)	(0.598)	(0.620)
School context (type of school)				
Municipal-public schools	0.934	0.768	-7.941*	-7.133*
	(5.822)	(2.673)	(3.109)	(3.109)
Subsidised-private schools	(Ref)	(Ref)	(Ref)	(Ref)
Private schools	9.204**	20.710***	8.448***	27.794***
	(2.501)	(1.779)	(1.870)	(1.915)
Students' characteristics				
Women	7.174***	-6.225***	-1.065	-5.362***
	(1.339)	(1.085)	(0.741)	(0.560)
SIMCE 4th grade		0.450***	0.153***	0.542***
		(0.007)	(0.016)	(0.006)
SIMCE 8th grade			0.866***	0.866***
			(0.024)	(0.024)
Constant	187.868***	172.643***	-43.859***	134.393***
	(6.580)	(7.436)	(6.890)	(5.626)
Number of students	37,000	37,000	37,000	37,000
Residual SIMCE ("Indirect effects")	-	Yes	No	Yes
R-squared	0.107	0.369	0.597	0.597

Notes: Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure A1. Direct and indirect effects of previous tests on current test score