Journal of Simulation

ISSN: 1747-7778 (Print) 1747-7786 (Online) Journal homepage: https://www.tandfonline.com/loi/tjsm20

An agent-based model of school choice with information asymmetries

Diego A. Díaz, Ana María Jiménez & Cristián Larroulet

JOURNAL OF SIMULATION
https://doi.org/10.1080/17477778.2019.1679674

THE OPERATIONAL RESEARCH SOCIETY
Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

An agent-based model of school choice with information asymmetries

Diego A. Díaz [a], Ana María Jiménez [b] and Cristián Larroulet [a]

[a]School of Business and Economics, Universidad del Desarrollo, Santiago, Chile; [b]School of Education, Universidad del Desarrollo, Santiago, Chile

ABSTRACT
Going from a neighbourhood-based to a choice-based system of school selection can have positive effects on enrolment in higher achievement schools, increasing average student achievement. We develop an Agent-Based model (ABM) that simulates students' decisions on a heterogeneous agents' framework with information asymmetries between income levels, allowing to simulate school choice policies and determine their impact on school enrolment and average student achievement. We use data from Santiago schools to initialise the model and study the impact of a discrete information signal of school achievement, as a policy implemented in 2010 in Chile called *traffic lights*.

ARTICLE HISTORY
Received 28 September 2018
Accepted 6 October 2019

KEYWORDS
Agent-based model; school choice; public policy; heterogeneous agents

1. Introduction

School choice, a term generally referring to K–12 public education options and studied substantially in the fields of education and economics, was relatively unheard of until Friedman wrote his seminal article in 1955, "The role of government in education" (Friedman, 1955). Back then the concept was not very important since students from public schools were, in most of the world, assigned to a school by the location of their family residence. More explicitly, school choice refers to an education system in which governments make direct or indirect payments to families that allow them to select a school, whether this school is privately or publicly managed. The objectives of school choice policies are to increase parental choice, promote school competition and allow low-income families access to private schools (Mizala, Romaguera, & Ostoic, 2005).

Policies, controlled experiments and surveys have been evaluated in different countries to determine the impact of school choice on specific outcomes, as student enrolment and test scores. Important examples are Andrabi, Das, and Khwaja (2017) and Hastings and Weinstein (2008), who performed experiments to study the degree to which the provision of information on school academic performance impacts parents' decisions of school selection. The results showed an increase on the fraction of parents selecting high performing schools and evidence that attending higher scoring schools increases students' test scores.

In an environment that allows for school choice, there seems to be a consensus among scholars that the provision of information has a positive effect on students' achievement levels. However, more research is needed regarding the mechanisms by which it impacts students' (or their parents) decisions regarding school selection and the school system. Our paper seeks to contribute to this area by simulating a schooling system with a dynamic Agent-Based Model (ABM) with a market of interacting schools and students. We follow Maroulis, Bakshy, Gomez, and Wilensky (2014), who developed an ABM to model the transition to public school choice in the Chicago public school system, implementing several characteristics from their model, as will be noted in section 3 when we describe the model in detail; however, we adopt a more general approach, allowing the model to apply to a broader range of different settings. A simulation approach is ideal for our objective because it requires to define the parameters and behaviour of every agent involved, allowing us to capture the dynamic processes involved.

Our model allows to simulate the application of two different types of policies in a context of utility maximising heterogeneous agents with information asymmetries: The first is a standard school choice policy, consisting on a shift from a neighbourhood-based to a school choice-based selection system, as in Maroulis et al. (2014). The second is a discrete information signal policy based on one that was implemented in Chile in 2010 called *traffic lights*, which consisted on giving simplified information of schools' achievement to households. For this reason, we calibrate the model to the capital city of Chile, Santiago, before the policy was applied, and compare the results afterwards, discussing the possible frictions that may cause our model to diverge from the data.

Our reasons for looking at the case of Chile and using the city of Santiago to set up our model and study the

CONTACT Diego A. Díaz diegodiaz@uchicago.edu School of Business and Economics, Universidad del Desarrollo, Santiago, Chile
Supplemental data for this article can be accessed here.

traffic lights, are three. First, Chile has had a large school choice system since 1981, which is the most extensive in the world in terms of the share of students affected (higher than 90%) (Gómez, Chumacero, & Paredes, 2012). Second, because of the scale of Chile's school choice program, the information signal policy *traffic lights* is the largest natural experiment in the subject in terms of share of students affected. Third, to our knowledge, this is the first article to study the *traffic lights*, although two graduate-level theses have been made in the subject, with different results (see Allende, 2012; Navarro-Palau, 2016).

The rest of the paper is organised as follows. In section 2 we describe the details of the schooling system in Chile as well as briefly discuss some of the most relevant observations about school choice, which we incorporate later in our ABM. At the end of section 2, we describe the *traffic lights* in detail. In section 3 we describe the model and our assumptions. In section 4 we explain the model calibration according to the data. In section 5 we present the results of the model and consider its implications as well as limitations. Finally, section 6 concludes and summarises our findings.

2. School choice in the Chilean school system

The case of Chile is interesting since a voucher school choice system was implemented in 1981, in which all parents became free to choose schools and voucher schools started receiving a subsidy for each student enrolled. After the introduction of the policy, and until today, every school in Chile falls into the following categories:

- Public: Administered by municipalities and financed primarily by vouchers paid by the government.
- Private-voucher, or subsidised: Administered privately and mainly funded by vouchers. It used to be allowed for this type of school to ask parents to pay an additional amount, but it was prohibited in 2016 in law number 20.845. From then on, all private-voucher schools are tuition free.
- Private: Administered privately and financed exclusively by parents and private donations.

The voucher system has been successful at the national level, or at least there is no evidence to say it has been unsuccessful when looking at aggregated educational indicators. It has increased competition by making every type of school compete with each other and it has increased enrolment since it provides higher incentives for schools to attract students (Larroulet & Gorosabel, 2015). Gallego (2006) shows that the higher competition brought about by the policy has improved the scores on national standardised tests in both public and subsidised schools. When looking at national indicators, Chile has done increasingly well on international standardised tests. Looking at the PISA report from 2015 (OECD, 2016), the country registered the best performance in all tests (science, reading, and mathematics) within the region (Latin America and the Caribbean). It also recorded the highest test inclusion in the area, reaching 80% of 15-year-olds.

There is evidence that competition has not been entirely effective, in the sense that low performing schools have not been forced out of the market by low enrolment (Contreras, Sepúlveda, & Bustos, 2010; Román & Perticará, 2011). Sapelli (2003) argues that Chile has not implemented a textbook case of the voucher educational system since no public schools have closed, and all the schools that lost students to private-voucher schools have received subsidies to pay their operational costs when needed. Given this problem, Sapelli continues, professors of public schools facing competition from subsidised schools have incentives for students to leave, since they keep their jobs, and get to teach smaller classes.

One of the factors contributing to the issue of low-performance schools staying in business is the information asymmetry between parents of different socioeconomic status. According to Elacqua and Fabrega (2004), the precision in the information that parents have about schools is a function of socioeconomic level. Another factor is the fact that school choice preferences are related to socioeconomic status (Thieme & Treviño, 2013). Hanushek, Kain, Rivkin, and Branch (2007), Hastings, Van Weelden, and Weinstein (2007) and Hastings, Kane, and Staiger (2005) argue that lower-income families give less importance to academic variables than higher socio-economic level families. This issue is relevant for our paper since it would reduce the effect of information-shock policies on low-income students. Regarding differences in school mobility (students changing schools), differences have also been found between socioeconomic levels (Larroulet, 2011).

Some authors have noticed that students of lower socioeconomic backgrounds can be concentrated in the lower performing schools when educated parents are the only ones that choose and demand more academic performance from schools (Berry, Jacob, & Levitt, 2000; Hsieh & Urquiola, 2006; Ladd, 2002; Ladd & Fiske, 2001; Mizala et al., 2005). This issue can be noticed in the Chilean educational system (see map of Santiago in the appendix), where socioeconomic status is correlated with school performance (Mizala, Romaguera, & Urquiola, 2007) and there is a relatively high level of segregation by socioeconomic level by communes.

2.1. *The* traffic lights *policy*

The *traffic lights* was put forward by the Chilean Department of Education to deliver general information about schools' standardised tests to families all over the country. Instead of giving a report with achievement scores by school as in the reports given in Hastings and Weinstein (2008) experiment, the Chilean government delivered a colour-coded map of each family's neighbourhood, with every school in the area marked in red, yellow, or green colour, depending on that school's performance on the previous year in a national standardised test called SIMCE. The performance was measured in terms of that school's students' average achievement in the test. A school with one standard deviation above average was marked green, a school with one standard deviation below was marked red, and all in between, yellow (Allende, 2012).

The reason for the policy's name, *traffic lights*, was straightforward; a red colour would signal to parents to stop their selection process and avoid the school since its standardised test scores are lower than average. A green colour would indicate to proceed and enrol your kids in the school since standardised test scores are higher than average, and yellow meant to proceed with caution since scores are average.

3. Model

We develop an ABM of school choice with a market of interacting schools and students. The reason for choosing Agent-Based Modelling as our methodology is that students interact with schools by choosing the one that maximises their utility. Since several attributes in the model are randomly selected, such as location, income and achievement, the aggregated outcomes are too complex to determine *a priori*, and we must resolve to model the individual agents' behaviour in order to study the emergent behaviour of the agents together. To our knowledge, this is the only ABM in the subject besides Maroulis et al. (2014). We follow Maroulis et al. (2014) but differ in that our simulation allows for heterogeneity and information asymmetries between income levels, allowing us to model an imperfect information shock as the *traffic lights* policy.

Our research also contributes to the literature of school choice by allowing a more general setup than Maroulis et al. (2014), which is adapted to the Chicago public school system in geography. Our setup is independent of location, with schools and students located randomly in the simulated environment. In the next section, we calibrate the model to Santiago, Chile. Parameter calibration consists of distributions of income and achievement, a parameter from the students' utility function, the share of the population classified as high-income, and the share of schools that are private.

3.1. Model description

The simulated environment is represented by a grid of dimensions 51×51 units with no boundaries,[1] populated by schools and students. The number of schools and students is chosen by the user,[2] with a location determined by a random draw from a uniform distribution of all positions on the grid. We use NetLogo 6.0.4 as the development environment, a free software package created by the Centre for Connected Learning and Computer-Based Modelling at Northwestern University by Uri Wilensky (1999).

Students in the simulation have the following attributes:

- {*xcor, ycor*}: Home location coordinates, or geographic location in the simulated environment. Drawn from a uniform distribution.
- *target*: Chosen school, the school that is selected by the students depending on their decision rules.
- *enrolled?*: Enrolment status. True or false variable indicating whether the student is currently enrolled.
- *income*: Income of the student, drawn from a Pareto distribution.
- *years-in-school*: Number of years in school remaining. Each student goes to school for 10 - periods[3] in total and graduates, at which point he disappears from the simulated environment.

Schools have the following attributes:

- {*xcor, ycor*}: School location coordinates in the simulated environment. Drawn from a uniform distribution.
- *enrolment*: Number of students enrolled in the school.
- *achievement*: Achievement of the school. Drawn from a normal distribution.
- *is-private?*: True or false variable indicating whether the school is private.

Schools may be either private or public, the difference being that only private schools charge tuition. Public schools can be government managed schools or private-voucher schools, which are privately managed but publicly funded, a very common type in the Chilean educational market.

A difference from Maroulis et al. (2014) is that, in their model, school achievement is determined by location and locations differ in socioeconomic status. In our setup, school achievement and location are drawn randomly and are not correlated. We do not have an attribute for student achievement, but instead, understand student achievement as the achievement of the school he is enrolled in.

We make several other assumptions about the market:

- Schools have no capacity constraints.
- Schools must accept all incoming students.

- When school choice is active, the model allows for the existence of private schools.
- There is no cost of tuition for public schools, but there is a *transportation cost*, which is proportional to the distance between the student and the school.
- Only private schools charge tuition.
- Schools can continue to operate regardless of how many students are enrolled in it.[4]
- The number of schools and their achievement level is fixed from the first period.
- High-income students have perfect information about school achievement.
- Low-income students have no information about school achievement but get a signal when the *traffic lights* policy is active.

All assumptions except the first two are sustained by the data or by articles about school choice that were previously cited. Assuming that schools have no capacity constraints and must accept all incoming students is a modelling simplification unlikely to have relevant repercussions to our conclusions. This is because as the law in Chile is set, public schools are not allowed to select students based on any factor such as previous or potential academic achievement; therefore, public schools, which compose most of the school system in Chile, accept all incoming students, except when limited by capacity constraints. Imposing a capacity constraint would likely influence the percentage of students enrolled in different achievement schools, but only in the short term. Considering public and private schools have incentives to enrol more students, it is likely that their capacity would expand quickly as a reaction to the increased demand.

The model allows for several user inputs. In the basic setup, without school choice, all students are required to attend a respective school in their neighbourhood, which is the closest one to their location (no schools are private in this setup, so there are no tuition costs). The simulation takes place in discrete steps, each representing a school year. Each period, the following actions occur:

(1) A new cohort of students enters the simulated environment.[5] Each student appears in a random location and must be in school until graduation (10 periods). The size of this cohort is 10 percent of the number of students so that the total number is the same every period – 10 percent in, 10 percent out.
(2) Each student chooses an enrols in a school, depending on whether there is school choice, and the *traffic lights* is on. The mechanism of this decision is the most important feature of the model and is described in detail in the next section. Students choose schools each period, but unless there is a policy change, they will not ever change their decisions since schools' achievement is set.
(3) A period is completed (or a school year), and students graduate when they complete 10 periods and exit the simulated environment.

In the visual representation of the model, we present schools as house-shaped figures on the grid, and students as human-shaped figures. We show each student's enrolment visually as a line between him and the school. Both types of agents are colour-coded, both for an easier visual understanding and for realistic modelling of the *traffic lights* policy since it consisted in giving maps to households with a colour designation of schools (green, yellow and red depending on their achievement level). We colour-code them the same way; we paint green those schools that are above one standard deviation from the expected value of achievement, we paint red those schools that are one standard deviation below, and yellow those in between. Private schools are colour-coded the same way but have a black roof. A screenshot for an example setup after one period is shown in Figure 1. The colour-coding of students is done in terms of income. The idea is to implement the observation reported in the literature that households make different decisions depending on income level, which is also an assumption made by the policymakers in Chile regarding the *traffic lights* policy, which is, that low-income individuals do not have, or have less information regarding school achievement, and therefore can benefit by the provision of information. Although we model the behaviour of students in our model, we can think of them as households, as they have an income and are responsible for the school choice decision. There is no distinction between these concepts in our model, so we use the terms "households" and "students" interchangeably.

On the simulated environment, we paint high-income students in blue, and low-income students in grey. The percentage of high-income students is selected by the user.

3.2. School choice mechanism

When school choice is allowed, students may choose a school for themselves. However, since low-income agents do not know the achievement of schools, this option is not available to them,[7] and they will choose the schools that are closest to them given that they can afford it (i.e. if the schools are publicly funded and they can cover the transportation costs).

Students maximise their utility according to a utility function depending on both achievement

Figure 1. Representative initialization of the model. A house figure represents a school, color-coded according to achievement level (green: high; yellow: medium; red: low). A person figure represents a student, color-coded according to income (blue: high; grey: low). Links (white lines) represent enrollment of a student in the school. The number in each house indicates the number of students enrolled in the school. A black roof indicates that the school is private[6] and students must pay tuition.

and proximity to the school as in Maroulis et al. (2014). However, an important difference in our model is that there is a transportation cost, therefore, agents are subject to a budget constraint. The problem they must solve is the following:

$$Max_i U(a_i, px_i) = a_i^{\alpha} px_i^{1-\alpha} \quad (1)$$

$$s.t. : I_i - Pd_i \geq 0$$

Where a_i is the achievement of school i, px_i is the student's proximity to school i, α is a parameter of the utility function that represents how much weight is given to achievement in comparison to proximity. I_i is the student's income, d_i is the Euclidean distance to the school and P is the school transportation cost per unit of distance. Proximity depends on the Euclidean distance and is normalised with respect to a maximum distance according to the following relation:

$$px_i = \frac{d_{max} - d_i}{d_{max}} \quad (2)$$

When agents maximise their utility, they will not necessarily choose the nearest school, unless such school is the only one that satisfies their budget constraint.

Tuition costs are not present in equation (1), which is because all private schools charge the same tuition, a value so that only high-income students may afford to enrol in them. As we will see in the next section, this idea is based on data from the Chilean school system and we chose not to incorporate tuition into the budget constraint to simplify the model.

To model the traffic lights policy, when the policy is turned-on, low-income students get a discrete signal of school achievement. The signal is the colour of the *traffic lights*, but agents cannot distinguish between the achievement of two schools of the same colour. Therefore, even though all students have the same utility function, low-income students' decisions will not necessarily be optimal, as they might be different than they would be with perfect information.

4. Data and model calibration

The databases used in our study[8] for model calibration are publicly available from the Ministry of Social Development, Ministry of Education (Mineduc) and *Agencia de Calidad de la Educación*. The educational data consists in school enrolment from the year 2006 until 2016, and the results of the standardised test score SIMCE, acronym in Spanish for Education Quality Measurement System, from the years 2006 until 2016. In total, in our data of Santiago, there is an average of 1768[9] schools between 2006–2016 and standardised test scores of such schools follow a normal distribution. Achievement in the model is modelled as a zero-truncated normal distribution with an expected value of 5 and a standard deviation of 1. The reason for centring the distribution in 5 is that the agents' utility functions become indefinite if achievement is negative.

Centring the distribution at 5 makes it highly unlikely that a school would have a negative achievement value, but it is truncated at zero to eliminate possible extreme values.[10] In the results section we re-centre the distribution at zero to facilitate the comparison with the data.

The SIMCE test is taken every year in Chile for students of 4th and 8th grade. Taking the average of maths and language scores of all students in a school, Figure 2 shows the distribution of Santiago schools' average achievement, which follows a normal distribution. The scores are standardised to have a distribution with zero mean and a standard deviation of 1. With data from the CASEN demographic survey, the income distribution of households is also presented, which follows roughly a Pareto distribution. The income distribution is, therefore, modelled as a Pareto distribution, with an alpha of 3.

There is a percentage of students that are not enrolled at school in Chile. The CASEN survey reports the main reason for this to be economical, as answered by 29% of survey takers (CASEN, 2011). This situation occurs despite having free public education options, and it is therefore modelled in our simulation as a student not being able to afford the school transportation cost. On 2011, a total of 86.701 children in school age were not enrolled in a school, out of a total of 3.496.879, which means that approximately 2.5% of children did not attend school that year (CASEN survey, 2011).

We define our base model as the parametrised simulation that replicates, on average, the equilibrium observed in terms of the aggregated enrolment variables (i.e. the share of students in high, medium, and low achievement schools) in Santiago before the *traffic lights* policy is implemented, that is, in 2009. We set a transportation cost so that there is a percentage of unenrolled students equal to 2.5%. We define an equilibrium as the average values of the aggregated enrolment variables once they settle to stable levels, which happens once the youngest generation of students who lived through a policy change, exit school (after 10 periods).

One of the main outputs of our model is the share of students that attend high (green), medium (yellow), and low (red) achievement schools according to the *traffic lights* classification. Figure 3 shows the percentages of students attending different achievement schools from 2006 to 2016. The α parameter from the utility function is adjusted with a Montecarlo simulation to produce, on average, an equilibrium equal to the one in 2009 according to the percentages observed in the data (68.0% in yellow coded schools, 19.5% in green coded schools, and 12.5% in red coded schools). Since we also consider the 2.5% of students that do not attend school, the real percentage of students in Santiago that attend different school achievement categories were: 67.2% for yellow schools, 18.7% for green schools, and 11.6% for red schools.

Before calibrating the value of α from the utility function, the percentage of high-income students must be selected. This parameter is important since it represents the share of agents that have perfect information about school achievement and can choose schools when school choice is available. In other words, this parameter models the information asymmetry that exists between households of different income categories. Even though in reality the information asymmetry most likely changes continuously among the income range, Schneider, Elacqua, and Buckley (2006) affirm that most empirical studies show that choosers are disproportionately higher-income than non-choosers. They also note that there are substantial differences in Chile by socioeconomic group on the process of selecting a school, observing that when dividing the socioeconomic groups in five, the two highest income groups choose almost only private schools and private voucher schools., while the bottom three groups only choose public and private voucher schools. Based

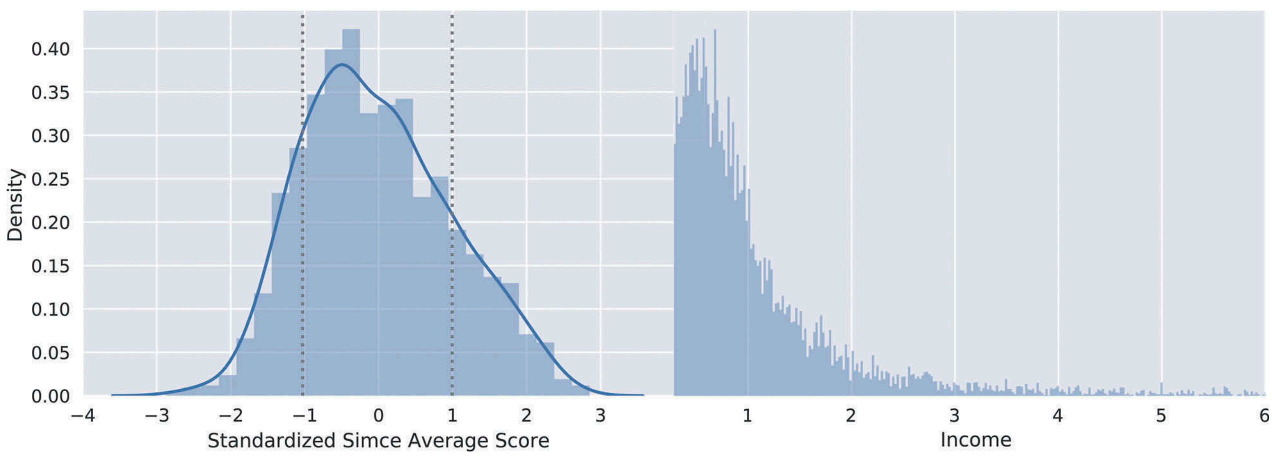

Figure 2. Distribution of average test scores by school, standardized to a normal distribution (0,1) on the left. Income distribution of households on the right, in millions of Chilean Pesos ($). Data for test scores from Mineduc and for income from the CASEN (2011) survey of the Ministry of Social Development of Chile. All data is of the city of Santiago.

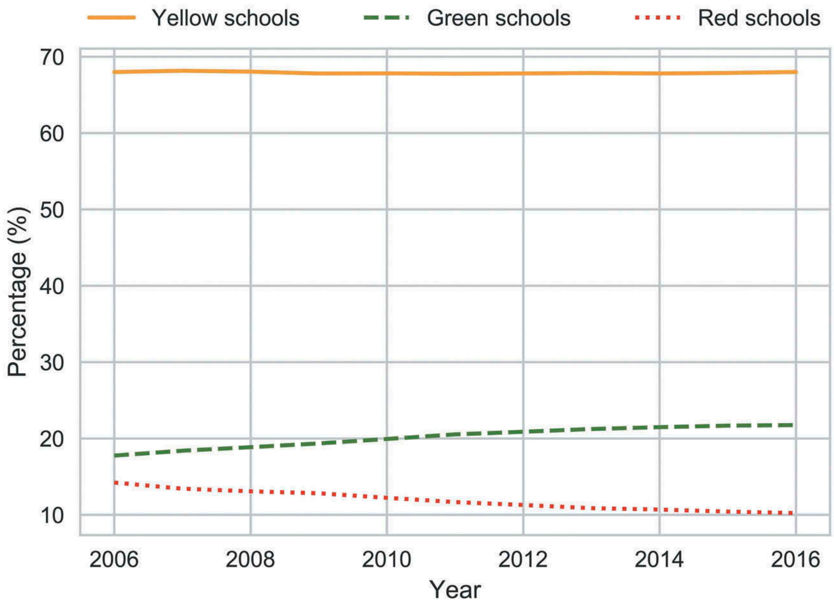

Figure 3. Percentage of students enrolled in schools by achievement level category according to the *traffic lights* policy. In our target year, 2009, these values were 68.0% for yellow schools, 19.5% for green schools, and 12.5% for red schools.

on Schneider et al. (2006), we consider the two highest income groups to calibrate the share of high-income students in our simulation. Since these groups account for 21 percent of total enrolment in the data, our variable takes this value.

The cost of tuition for private schools is also based on Schneider et al. (2006). Since they observe that only the two highest income groups attend private schools in Chile, we calibrate the model so that the result is the same by simply not allowing low-income agents attend private schools.

The percentage of schools that are private is calibrated directly from the educational data, according to their achievement categories. In 2009, 45.4% schools were private out of all the schools in the high achievement group, while in the medium achievement category, the percentage is 2.6%, and only 0.4% in the low achievement group. We randomly establish which schools are private according to probabilities given by these percentages.

With the share of high-income students set we can adjust the α from the utility function to produce the desired equilibrium before the *traffic lights* policy was implemented, obtaining a value of 0.25. Leading to the final calibration of the model, we performed multiple simulations with different parameter configurations to determine the values of α and the transportation cost, since these two are the only parameters that are not set from the data and must be adjusted to produce the enrolment aggregates observed in Santiago. We show some results of this process in the next section as a robustness test.

One of the limitations of our model is that the schools' locations are not calibrated to the geography of Santiago. This decision was made on purpose since we attempt to contribute to the theory of school choice, and therefore study the problem on a general setup. There is a trade-off between achieving this objective and providing more realism in the model to study the implications of an information signal as the *traffic lights* as it was implemented in Santiago. Although the locations of schools are not important for the policy to work, the distribution of achievement and income is, since having a not-random distribution of school achievement through the city and a not-random distribution of households' income will lead to different aggregated enrolment variables. The city of Santiago is relatively segregated in terms of both households' income and schools' achievement. We explain how this observation might be a relevant factor in explaining the differences between reality and our simulation in the next section.

5. Results and discussion

This paper models some of the dynamic processes that govern the school choice decision every student (or household) makes. By calibrating the model to the Chilean education system and the data from its capital city, Santiago, the model reaches an equilibrium in line with the aggregated enrolment variables observed before the *traffic lights* policy was implemented. Our set up of the model, including the calibrated parameters, is shown in Table 1.

After the setup, the model can be put in motion with the "go" button. Two figures are plotted that show the main results while the simulation is running. The first one shows the variables that are directly comparable with the data. Figure 4 shows the right side of the screen, which contains both the results and the distribution of the random attributes of each agent type,

Table 1. Model calibration of main specification.

Model variable	Value	Calibration reference
Number of schools	200	Scale variable*
Students	3000	Scale variable*
High-income student percentage	21%	Share of the two highest income groups (Schneider et al., 2006)
Utility function's α	0.25	Calibrated to match the 2009's aggregated enrolment variables.
School transportation cost	30	Calibrated to account for the 2.5% of unenrolled students. Data from CASEN (2011)
Pareto distribution alpha	3	Fitted to households' income distribution. Data from CASEN (2011)
Pareto distribution minimum	1	Fitted to households' income distribution. Data from CASEN (2011)
Probability that green school is private	45.40%	Share of green schools that are private. Data from Mineduc (2018)
Probability that yellow school is private	2.60%	Share of yellow schools that are private. Data from Mineduc (2018)
Probability that red school is private	0.40%	Share of red schools that are private. Data from Mineduc (2018)
School choice	On	Policy situation in Chile in 2009
Traffic lights policy	Off	Policy situation in Chile in 2009

*Relatively high values of the scale variables (number of schools and students) are needed to match the income and achievement distributions with the data. Scaling is necessary due to computer processing limitations and the Netlogo environment, however, the observed distributions are replicated. All other parameters are calibrated to directly match the data for the city of Santiago in 2009, before the *traffic lights* were implemented.

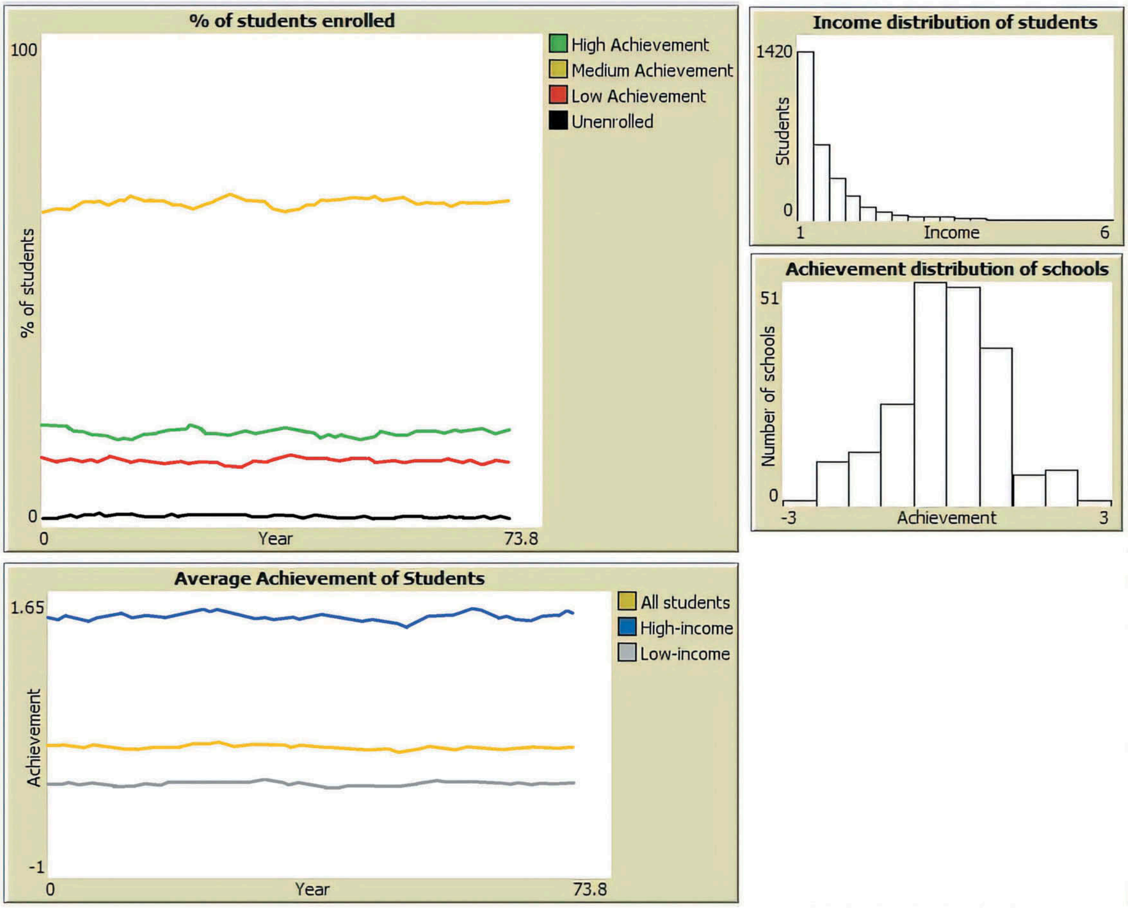

Figure 4. Output of our base model specification from Table 1. Upper left: share of students enrolled in each achievement category and share of unenrolled students; Bottom left: average achievement of students, total and by income category; Upper right: histogram of the income distribution of students, which follows a Pareto distribution; Bottom right: histogram of the achievement distribution of schools, which is normally distributed.

income for students and achievement for schools. Note that achievement is centred back at zero to facilitate comparisons.

In our ABM, turning school choice on, allows high-income agents to start maximising their utility function subject to their budget constraint, which can be immediately noticed in the model's output. Turning on the *traffic lights* policy gives a discrete information signal to low-income agents, allowing them to distinguish between school colours and maximise the same utility function, subject to their budget constraint, which is more restrictive since their income is lower. Figure 5 shows the model calibrated to Santiago with the policies starting at different times.

In the ABM without school choice the results are straightforward. The students must go to their assigned school, which is the nearest one to their location. On average between multiple simulations,

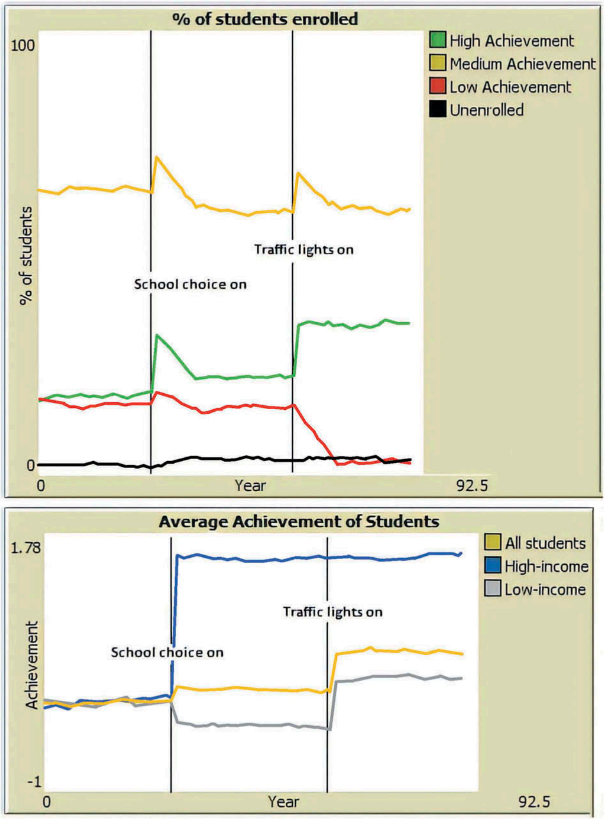

Figure 5. Output of running model calibrated to Santiago with both policies being turned on at different times. Above: Share of students enrolled by achievement category. Bottom: Average achievement of enrolled students by income level.

there are no differences in the share of students enrolled in high or low achievement schools on this setup, and any differences observed are due to differences in the random initial assignment of school locations.

Turning on school choice increases the share of students in higher achievement schools, which overshoots the following periods because of some students changing schools before they graduate. This may only happen on the years a policy is put in effect since from then on, the students will be facing the same problem until there is another policy change. The model settles to a level with fewer students enrolled in medium achievement schools, more enrolled in higher achievement schools, and relatively less in lower achievement schools.

When turning the *traffic lights* policy on, a much higher share of students start choosing higher achievement schools (38%), and a much lower share start choosing lower achievement schools (2.3%), which becomes even less than the share of unenrolled students. In Santiago, the results of the *traffic lights*, that is, the effect it had on the enrolment decisions of current and new students, was not as pronounced as our simulation predicts. As we show on Figure 3, the share of students enrolled in low achievement schools was never less than 10%.[11] We suggest the difference with the simulation occurs mainly because of two reasons, the first is an observation that our model does not incorporate, which is the geographical segregation of schools based on socio-economic level. From the data obtained from the Ministry of Education, which contains schools' location coordinates, we make a map of Santiago indicating the locations of schools by achievement categories (see Appendix Figure A3). The comparison between this map and our populated model is shown in the appendix, and the conclusion is that the location of schools in Santiago is correlated with school achievement. A local resident will notice that higher income neighbourhoods have a higher share of high achievement schools than low-income neighbourhoods. The distribution of schools may cause that some low-income households are not able to choose a higher achievement alternative even when they have information about school achievement from the *traffic lights*. A model that incorporates a mechanism for schools to choose their location based on demand would be an improvement to our model.

The second reason for the difference between our model and the data, is the fact that low-income families give lower importance to achievement variables, as noted by Hanushek et al. (2007), Hastings et al. (2007), and Hastings et al. (2005). A simple way to incorporate this observation would be to make the α from the utility function (share of utility derived from achievement), depend on the student's income level.

Going back to our results, we observe that without school choice, the average achievement of students is zero, which is expected since location and achievement are randomly assigned. Therefore, average achievement of students equals the average achievement of all schools. When school choice is turned on, high-income students will start enrolling in higher achievement schools, which increases their average achievement to over 1.5 standard deviations above the mean. Considering all students, the net effect of school choice in average achievement is positive, both in the model and in the data. In the simulation, low-income students start enrolling, on average, in slightly lower achievement schools than before having school choice. This occurs because some schools become private and are unavailable to low-income students. The average achievement level of low-income students is below the mean, as is in the data. Figure 6 shows average achievement of all students and of *priority* (very low-income) students on Santiago from 2008 to 2016.

Priority students is a designation used by the government of Chile for low-income students that get a 50% higher value voucher for education than regular students.[12] The distinction is based on income level, so the lower income students are the ones who get the benefit. We show results for priority students because the educational data does not specify income level. However, priority students make a substantial share of the total number of students (on average 30.9% of enrolment from 2008 to 2016 and increasing). In our calibrated model, 79% of students are classified as low-income, so although the comparison is not direct, it is a relevant point of reference.

The ABM predicts a higher increase on average achievement (0.5 standard deviations) from the *traffic lights* policy than what is observed on the data (0.13 standard deviations from 2009 to 2016). We attribute the difference to the same reasons previously mentioned for the aggregated enrolment variables, geographic segregation between income levels.

When the *traffic lights* policy is activated, the average achievement of high-income students does not change, which is expected considering that they have perfect information about school achievement. The average achievement of low-income students increases to 0.6 standard deviations above the mean. In the data, although the groups are not directly comparable, priority students increase their average achievement in 0.2 standard deviations between 2009 and 2016.

As a robustness test of the results obtained from our calibrated model, we include a sensitivity analysis regarding different parameter configurations. This includes simulations with different values of the parameter α from the student's utility function and the transportation cost. Figure 7 shows the aggregated enrolment variables by achievement category and Figure 8 shows the aggregated achievement variables by income level, with different values of α.

The model's sensitivity to α decreases for higher values of the parameter. As can be seen in Figure 7, for α = 0.3 and α = 1, the aggregated enrolment in different categories of schools are relatively similar, which happens because students are subject to a budget constraint, and since most students are low-income, their budget constraint is more restrictive. After the *traffic lights* policy is applied, enrolment in lower achievement schools (red schools) decreases slightly more when α = 1 than when α = 0.3, while enrolment in medium achievement schools (yellow) increases more. This effect is driven by low-income students since they are the only ones that might change their behaviour after the *traffic lights* policy. With α = 1,

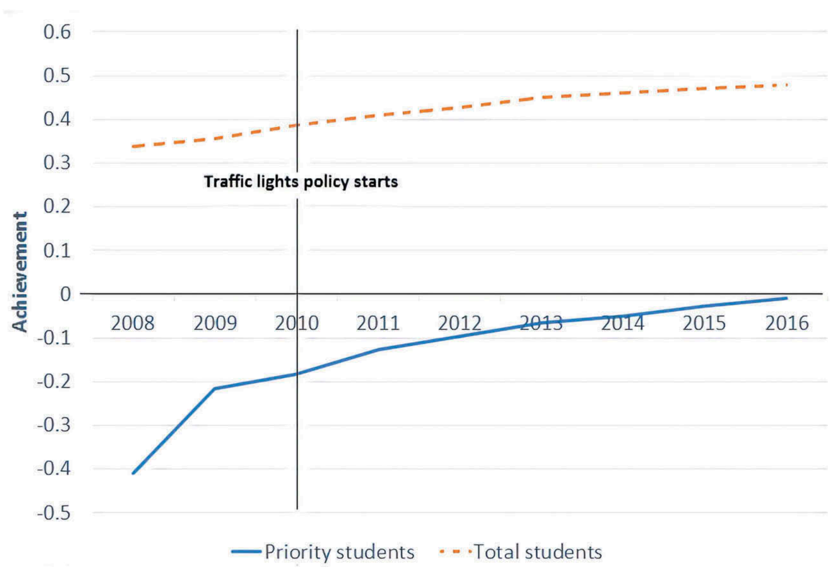

Figure 6. Average achievement of schools in which students are enrolled in. Total students in orange (upper line), and priority students in blue (lower line). Data: Mineduc (2018).

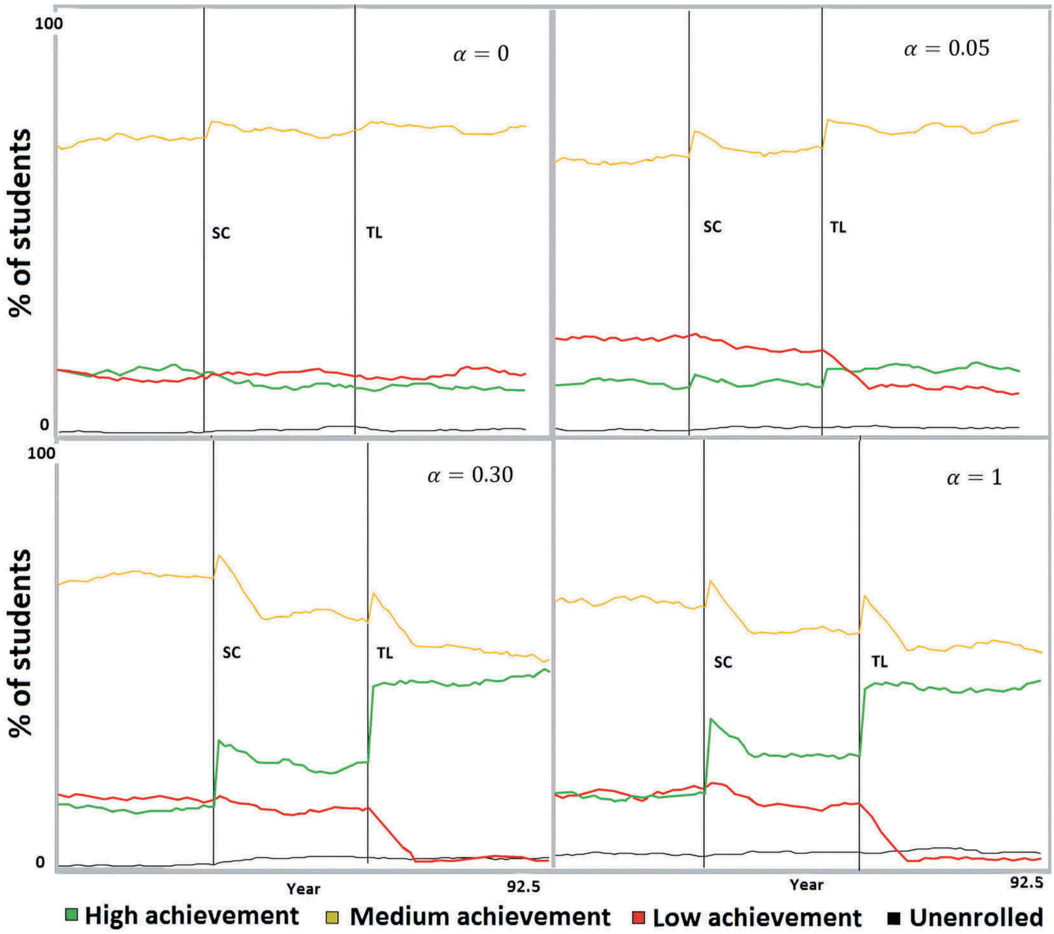

Figure 7. Aggregated enrollment by different values of alpha under different policy environments. The vertical black lines show the period in which there is a policy change. SC: School choice is activated. TL: *Traffic lights* policy is activated. All other parameters stay the same as in Table 1.

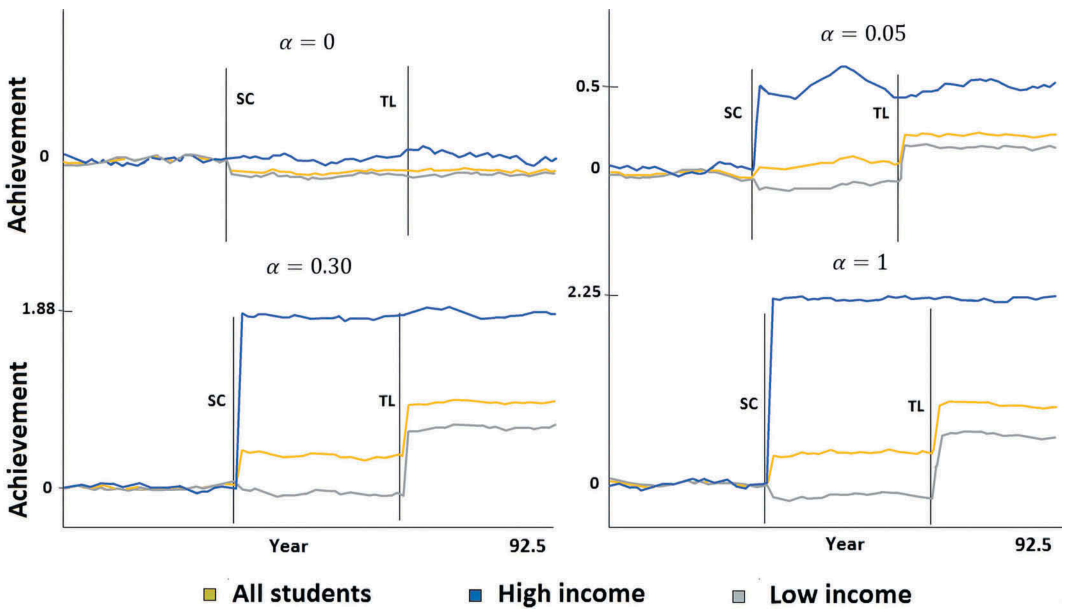

Figure 8. Average achievement by income category by different values of alpha under different policy environments. The vertical black lines show the period in which there is a policy change. SC: School choice is activated. TL: *Traffic lights* policy is activated. Top-left: α = 0. Top-right: α = 0.05. Bottom-left: α = 0.30. Bottom-right: α = 1. All other parameters are the same as in Table 1.

students only value achievement and do not consider distance; therefore, they will attend the highest achievement school that they can afford.

Looking at Figure 8, we can see how the achievement level of low-income students increases after the *traffic lights* policy is applied for all values of α > 0. Note that when they do not value achievement (α = 0), giving them information about it will not alter their behaviour. The increase in average achievement of low-income students is higher when the value of α increases but up until a certain point. Negligible differences can be observed between the cases when α = 0.3 and α = 1 because not many of them can afford to go to higher achievement schools at that point. Regarding high-income students, when α > 0, they start enrolling more in higher achievement schools when shifting from a neighbourhood-based to a choice-based school system. The span of this variation goes from an average achievement of zero when they do not value achievement, to around 0.5 standard deviations over the expected value of achievement with α = 0.05, up until 2.25 standard deviations above when α = 1.

On Figures 9 and 10 we change the transportation cost instead of α and keep all other parameters the same as in our base model. When increasing the transportation cost, two things happen: first, fewer students can attend school, increasing the share of students that are unenrolled. Second, the pool of available schools to each student is reduced because the transportation cost is a component in their budget constraints. This makes policies less effective as tools for increasing enrolment in higher achievement schools. When transportation costs are at the maximum allowed (TC = 100), the effect of the policies in enrolment aggregates is almost imperceptible.

The effects of increasing transportation costs affect both income categories. Looking at Figure 10 we can see the effect on both groups. Changing the transportation costs from 66 to 100 produces almost no effect on low-income students since their choice set is extremely limited at that point and almost all of them are not enrolled. For high-income students, the increase in average achievement when turning on school choice becomes smaller when the transportation cost is higher.

Simulation methods have countless applications in multiple fields, as they allow researchers to capture the dynamic processes involved and determine a system's equilibrium by allowing the agents interact with each

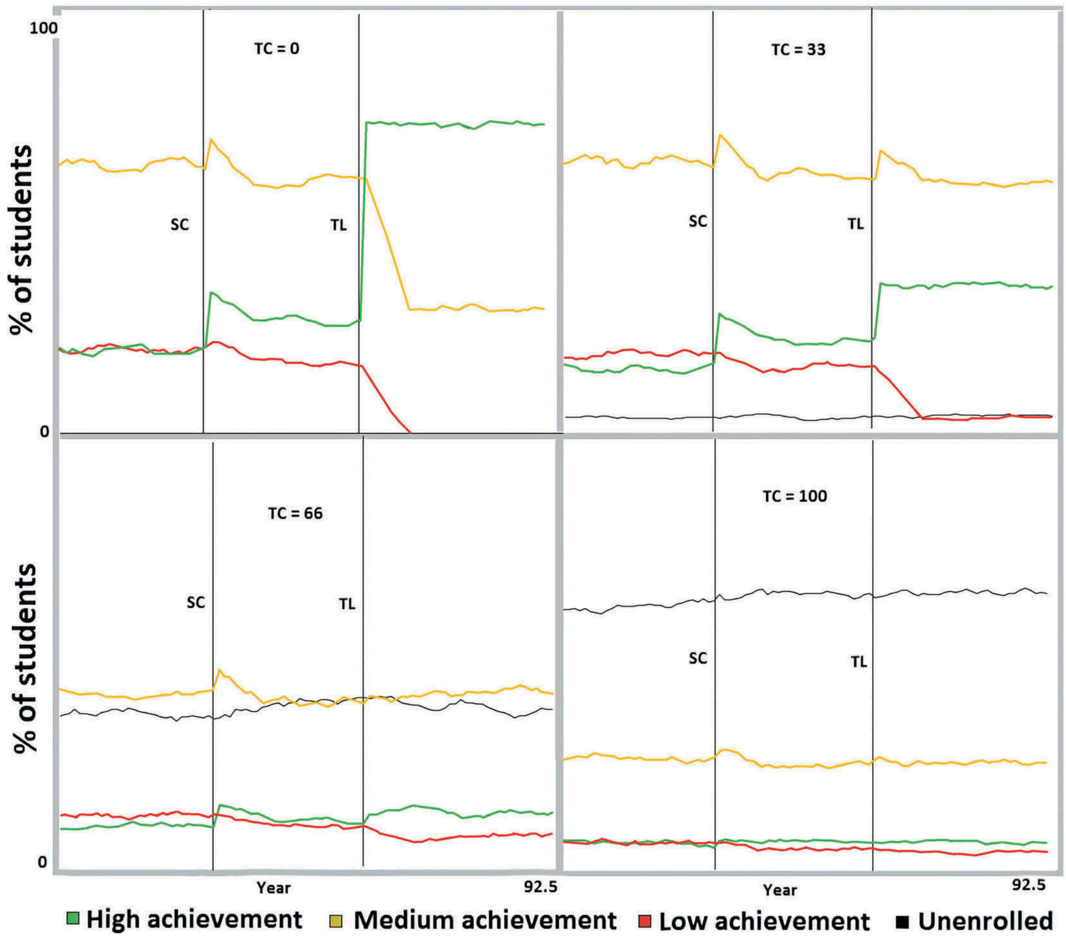

Figure 9. Aggregated enrollment by different values of transportation cost at different policy environments. The vertical black lines show the period in which there is a policy change. TC: Transportation cost. SC: School choice is activated. TL: *Traffic lights* policy is activated. Top-left: TC = 0. Top-right: TC = 33. Bottom-left: TC = 66. Bottom-right: TC = 100. All other parameters are the same as in Table 1.

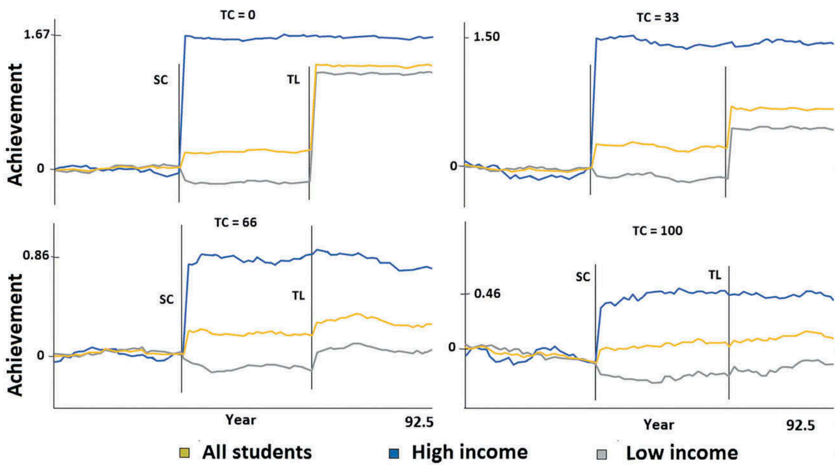

Figure 10. Average achievement by income category with different values of transportation cost under different policy environments. The vertical black lines show the period in which there is a policy change. TC: Transportation cost. SC: School choice is activated. TL: *Traffic lights* policy is activated. Top-left: TC = 0. Top-right: TC = 33. Bottom-left: TC = 66. Bottom-right: TC = 100. All other parameters are the same as in Table 1.

other in a simulated environment. We defined a theoretical school market and simulated the behaviour of students as well as their interactions with schools. By initialising the model with data from Santiago, our model has shed light on the mechanisms involved in the students' decision process and allows researchers and policy makers to assess the impact of two education policies.

6. Conclusion

On this paper, we developed an ABM of school choice with heterogeneous students and schools to understand the main behavioural processes that govern the school selection decision. Our simulation allows us to assess the impact of going from a neighbourhood-based to a choice-based system of school selection and of a policy that gives a discrete information signal of school achievement to households. We found that establishing a choice-based system increases aggregated enrolment in higher achievement schools, driven by the increase of enrolment of students that have information about school achievement. Extending from Maroulis et al. (2014), we identified and incorporated three features that are relevant in the school selection process: an asymmetry of information between students' income levels, a budget constraint in the students' utility maximisation problem, and a transportation cost.

By looking at the data of Santiago and calibrating our ABM to the city's enrolment aggregates and households' income distribution, we found that the provision of information by the *traffic lights* policy increased aggregated average achievement by increasing the average achievement of low-income students. From a policy perspective, our work is relevant for two reasons. First, it validates the benefit of information shock policies as the *traffic lights* for improving the school selection decision of low-income households. Second, it allows users to simulate the school system and study policies in a simulated environment, obtaining a prediction of aggregated enrolment variables.

By simulating the *traffic lights* policy, our model predicts an increase in enrolment of low-income students in higher achievement schools, but on a larger magnitude than what is observed on the data for Santiago. We suggest the difference occurs for two reasons. First, due to the lack of geographical segregation of schools based on school achievement, a characteristic observed on the distribution of schools in Santiago, which decreases the set of available higher achievement schools to lower-income students. Second, to the fact that households with different income levels could have different preferences regarding school achievement. This last idea derives from the observation that some authors have noted, that low-income families place less weight on academics than high-income families (Hastings et al., 2005).

Future lines of research could incorporate mechanics to account for higher geographical realism, segregation dynamics, and different preferences between income groups, for example, by allowing schools to choose locations based on demand and by incorporating income into the students' utility function.

Notes

1. The simulated environment has no boundaries to avoid differences in behaviour from the agents located near the edges. This way an agent located near the edge can cross through the edge to the other side of the simulated environment. A boundless setup is different from Maroulis et al. (2014), which represents the Chicago geography in a grid of dimensions 18×18.
2. In Maroulis et al.'s (2014) model, the numbers of students and schools are fixed; we allow user input for the values of these variables.

3. We set the number of periods that a student has to attend school to 10. We can think of the total period as pre, primary and middle school combined. The length of the period can be adjusted depending on location or the type of education studied in the model settings.
4. Román and Perticará (2011) and Contreras et al. (2010) argue that low performing schools have not been forced out of the market by increased competition in the Chilean educational market.
5. At setup, to start the model, most students are not part of the new cohort, but are created with a random value for the attribute years-in-school, drawn from a uniform distribution of natural numbers from 1 to 9, some will graduate in the first period, and some will be in school for 9 more years, but none of them will have 10 years remaining as the new cohort.
6. The model is like the Chilean education system, in which there are private schools that charge tuition and private-voucher schools that do not charge tuition. Private-voucher schools are privately managed but publicly funded. In the model, all public schools behave the same way, so they represent both public schools and private-voucher schools.
7. The difference between high and low-income agents besides income, is that those that are high-income have perfect information about school achievement and those who are low-income have no information (unless the *traffic lights* is in effect). To give all agents the ability to choose with perfect information, one can simply set the parameter "percent of agents as high-income" to 100%. The share of high-income students can also be understood as the percentage of choosers, as in Maroulis et al. (2014).
8. The model and all data used in this paper is available for download as supplemental material.
9. The value of 1768 schools in Santiago comes from the Mineduc (2018) database considering a balanced panel. It only includes schools that have complete enrolment data between 2006 and 2016.
10. By truncating the distribution in zero we are saying that zero is the lowest achievement possible a school might have, and by the way the agents' utility function is defined, such a school would give the minimum possible utility.
11. The data in Figure 3 does not consider the share of students that do not attend school. If it did, the share of students that attend low-achievement schools would be slightly smaller.
12. The concept of priority student was created in 2008 by the government of Chile with law number 20,248 of Subvención escolar preferencial (priority school subsidy).

Acknowledgments

We are thankful for useful discussions with Juan Pablo Couyoumdjian, Daniel Lerner, Constanza Gomez, Victor Martinez and Zimu Wan. We are also grateful to Andrés Marticorena for excellent research assistance. Any error is the authors' fault alone.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Universidad del Desarrollo with funding from Concurso Interfacultades.

ORCID

Diego A. Díaz http://orcid.org/0000-0002-5897-5340

References

Allende, C. (2012). *The impact of information on academic achievement and school, choice: Evidence from Chilean traffic lights* (Master's Dissertation). Pontificia Universidad, Catolica de Chile.

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, *107*(6), 1535–1563.

Barro, R., & Lee, J. W. (2011). Barro-Lee educational attainment dataset. Retrieved from BarroLee.Com

Berry, J., Jacob, B., & Levitt, S. (2000). The impact of school choice on student outcomes: An analysis of the Chicago public schools. NBER Working Paper No 7888.

CASEN survey. (2011). Ministry of social development, social division. Retrieved from www.ministeriodesarrollosocial.gob.cl/observatorio/casen/

Contreras, D., Sepúlveda, P., & Bustos, S. (2010). When the schools are the ones that chose: The effects of screening in Chile. *Social Science Quarterly*, *91*(5), 1350–1368.

Elacqua, G., & Fabrega, R. (2004). *El consumidor de la educación: El actor olvidado de la libre elección de escuelas en Chile*. Santiago de Chile: PREAL.

Friedman, M. (1955). The role of government in education. Economics and the public interest. University of Rutgers Press.

Gallego, F. (2006). Voucher-school competition, incentives, and outcomes: Evidence from Chile. Manuscrito, Instituto de Economía, PUC, 2006.

Gómez, D., Chumacero, R. A., & Paredes, R. D. (2012). School choice and information. *Estudios De Economía*, *39*(2), 143–157.

Hanushek, E., Kain, J. F., Rivkin, S., & Branch, G. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics*, *91*(5–6), 823–848.

Hastings, J. S., Kane, T., & Staiger, D. (2005). Parental preferences and school competition: Evidence from a public school choice program(No. 11805). National Bureau of Economic Research.

Hastings, J. S., Van Weelden, R., & Weinstein, J. (2007). *Preferences, information, and parental choice behavior in public school choice* (No. w12995). National Bureau of Economic Research.

Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, *123*(4), 1373–1414.

Hsieh, C. T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, *90*(8–9), 1477–1503.

Ladd, H. (2002). School Vouchers: A critical review. *Journal of Economic Perspectives*, *16*(4), 3–24.

Ladd, H., & Fiske, E. (2001). The Uneven paying field of school choice: Evidence from New Zealand. *Journal of Policy Analysis and Management*, *20*(1), 43–64.

Larroulet, C. (2011). *Análisis de la movilidad escolar en Chile* (Tesis de magíster). Pontifica Universidad Católica de Chile, Santiago, Chile).

Larroulet, C., & Gorosabel, J. (2015). La Educación en la Encrucijada. ¿Estado Docente o Sociedad Docente? 89. Ril editores.

Maroulis, S., Bakshy, E., Gomez, L., & Wilensky, U. (2014). Modeling the transition to public school choice. *Journal of Artificial Societies and Social Simulation*, *17*(2), 3.

Mineduc. (2018). Ministry of education of Chile. Enrollment at: http://datos.mineduc.cl/dashboards/19758/bases-de-datos-resumen-de-matricula-por-ee/ Achievement at: http://informacionestadistica.agenciaeducacion.cl/#/bases

Mizala, A., Romaguera, P., & Ostoic, C. (2005). Equity and achievement in the Chilean school choice system. Centro de Economía Aplicada. Departamento de Ingeniería Industrial. Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile.

Mizala, A., Romaguera, P., & Urquiola, M. (2007). Socioeconomic status or noise? Tradeoffs in the generation of school quality information. *Journal of Development Economics*, *84*(1), 61–75.

Navarro-Palau, P. (2016). The effects of distributing school performance information: Evidence from the Chilean "*Traffic Lights*". Chapter 2 of PhD thesis at Columbia University. Essays in Education Policies in Latin America. Retrieved from https://academiccommons.columbia.edu/catalog/ac:201039

OECD. (2016). PISA 2015 results (Volume I): Excellence and equity in education, PISA, Author, Paris. doi:10.1787/9789264266490-en

Román, M., & Perticará, M. (2011). La elección y permanencia en colegios de baja calidad en Chile. *Revista Uruguaya De Ciencia Política*, *21*(1), 43–65.

Sapelli, C. (2003). The Chilean Voucher System: Some new results and research challenges. *Cuadernos De Economía*, *40* (121), 530–538.

Schneider, M., Elacqua, G., & Buckley, J. (2006). School choice in Chile: Is it class or the classroom? *Journal of Policy Analysis and Management: the Journal of the Association for Public Policy Analysis and Management*, *25*(3), 577–601.

Thieme, C., & Treviño, E. (2013). School choice and market imperfections: Evidence from Chile. *Education and Urban Society*, *45*(6), 635–657.

Wilensky, U. (1999). NetLogo. [computer software]. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL. Retrieved from http://ccl.northwestern.edu/netlogo

Appendix

Figure A1 shows a map of a neighbourhood in Santiago as was delivered to families in the same neighbourhood as part of the *traffic lights* policy.

Figure A2 shows the average years of education for countries in Latin America according to the Barro and Lee (2011) dataset. Chile, which drastically changed its public education system in the early 1980s, now leads the region in terms of this variable.

Traffic lights colour designation

Since data for the actual classification of colours of schools according to the *traffic lights* policy is not available (Navarro-Palau, 2016), the classification must be performed by ourselves according to the decision rule. Allende (2012) shows the decision rule required for replicating the government designation.

The databases used in our study are from Mineduc and Agencia de Calidad de la Educación, both Chilean government institutions managed by the Ministry of Education (Mineduc). The data consists of school enrolment from the year 2006 until 2016, and the results of the standardised test score SIMCE, acronym in Spanish for Education Quality Measurement System, from the years 2006 until 2016.

Regarding the policy, the idea of Mineduc in 2010 was to assign one of three colours to schools in order to inform those colours to parents so they might choose green schools and avoid, or move away, from red ones. The method to determine the colours was as follows: Four Simce scores are taken from the year 2009 for each school, which are maths and language scores for both 4th and 8th grade. These scores are standardised with the mean and standard deviation from the distribution of all schools to a standard normal distribution. However, not every school has data reported for each of the two grades, which can happen because the school is only a primary school, or it is a secondary only school, or had too few students taking the test, or simple because the data was lost or not measured. To account for this issue, schools are classified into three types:

Type 1: Schools with only 4th grade scores.
Type 2: Schools with both 4th and 8th grade scores.
Type 3: Schools with only 8th grade scores.

The rule of colour assignment, implemented within each school type, is as follows:
Red schools: Schools with a score lower than one standard deviation from the mean of the distribution.
Yellow schools: Schools with a score that falls within one standard deviation from the mean of the distribution.

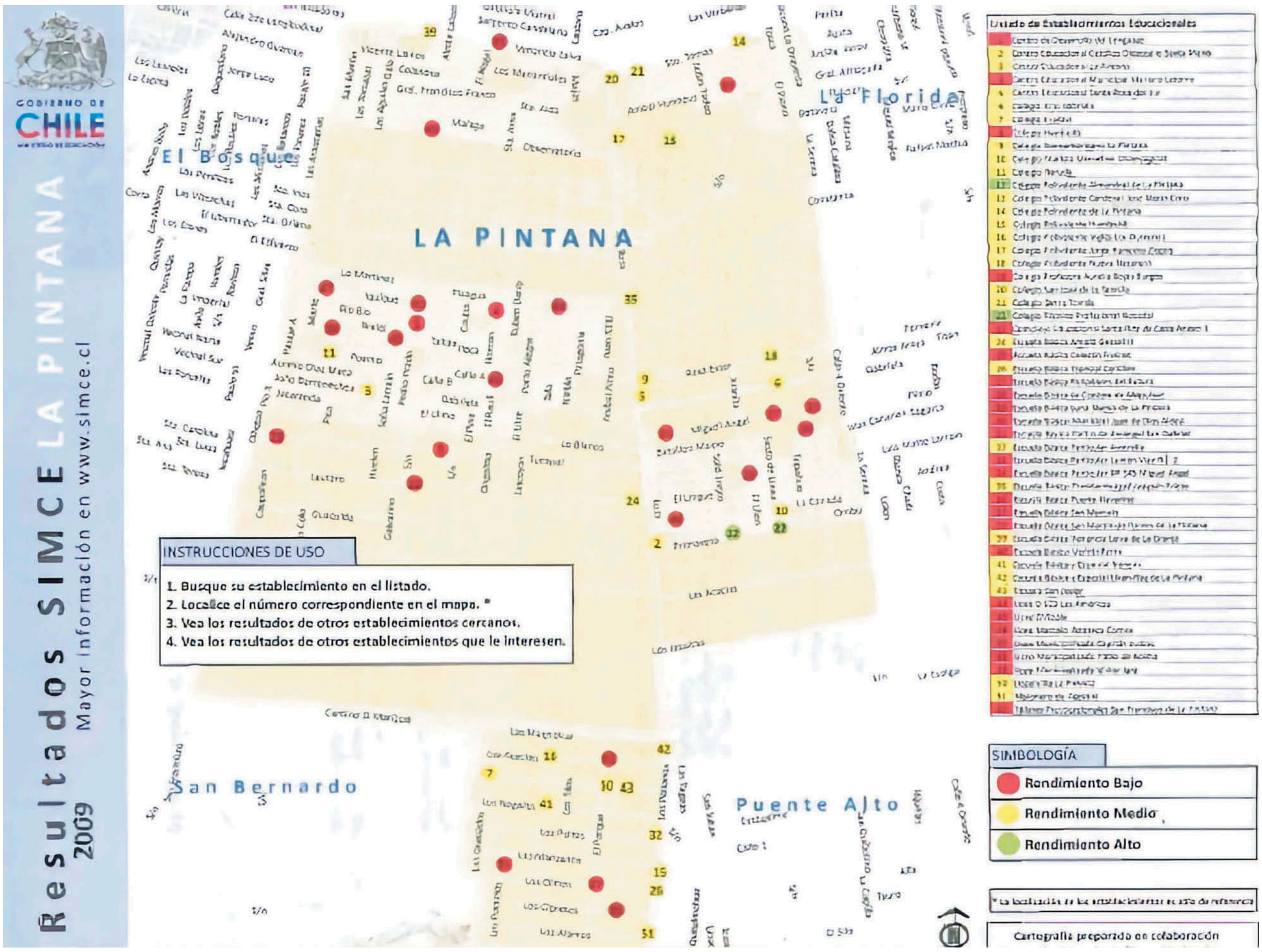

Figure A1. Map as given to parents in La Pintana neighborhood in Santiago. On the center: map with the schools marked on it with a number identifier. On the upper right: name of every school in the neighborhood with its corresponding number and color from the *traffic lights*. Bottom right corner: description of the color denomination.

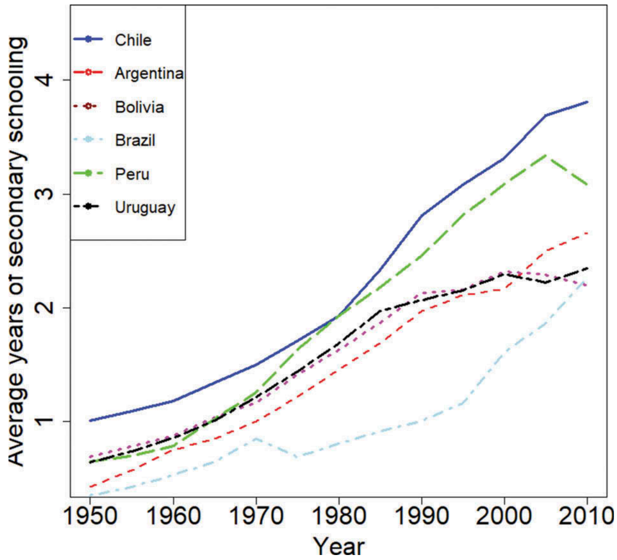

Figure A2. Years of secondary schooling in Latin America. Data: Barro and Lee (2011).

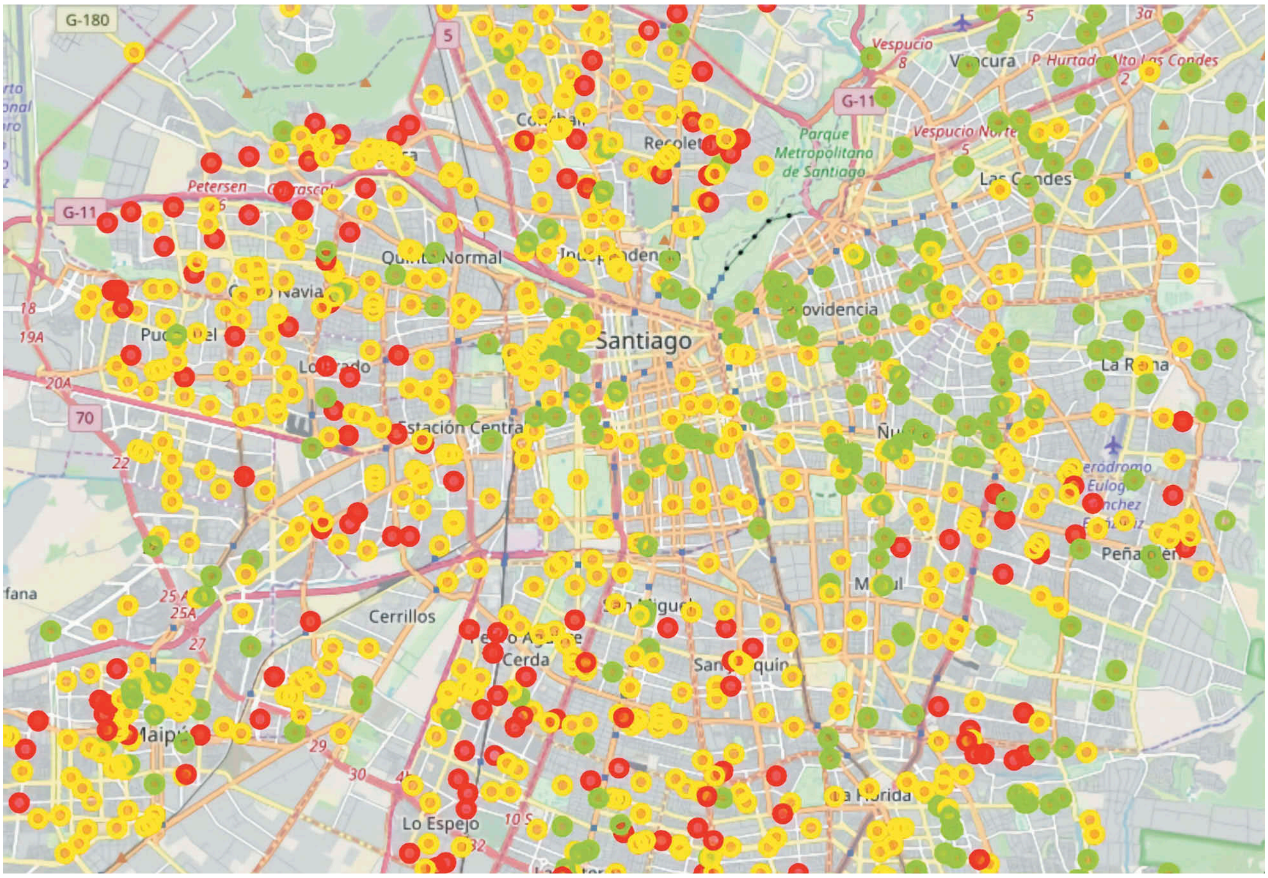

Figure A3. Map of Santiago with schools marked according to the traffic lights designation as it was estimated. Data: Mineduc (2018).

Green schools: Schools with a score higher than one standard deviation from the mean of the distribution.

Since type 2 schools have two scores for making the comparison, the average is taken between 4th and 8th grade.

It is relevant to add that since the policy was put in effect near the middle of the school year on June 2010, the effects should not be noticed until 2011, since the school year starts in March and therefore all new students that might be affected by the policy would start school from 2011. Regarding the students that change schools at the middle of the year, this is not a common decision, especially considering that the spring semester starts at the beginning of august, so it would be unlikely that the policy would be responsible for the change. It is reasonable to assume that more frequently, this type of behaviour has to do with personal reasons than wanting to be in a higher performing school.

Figure A3 presents a map of Santiago with the colour designation of schools. As can be seen, there are sectors with very different densities of schools by achievement level. Higher achievement is concentrated in the centre and north east, while most of the other sectors have a higher concentration of medium and low achievement schools.

To assess that segregation exists in school achievement by neighbourhood in Santiago, the percentage of schools by achievement category against average neighbourhood family income was calculated and is shown in Figure A4.

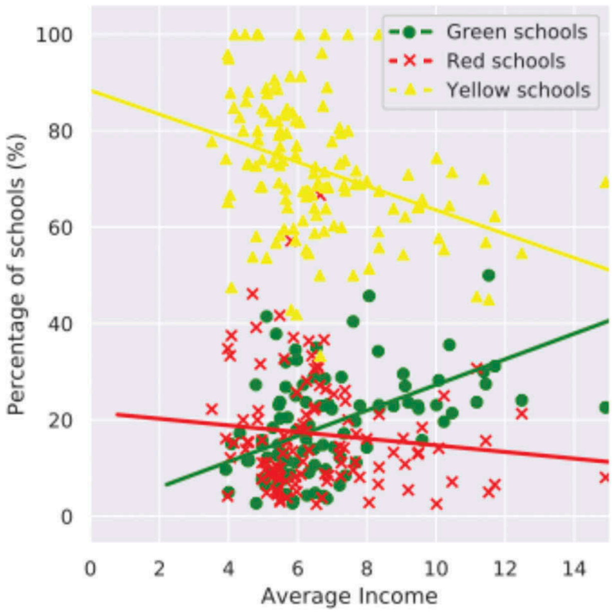

Figure A4. Percentage of schools by achievement category by commune's average family income. Income values in 10^5 Chilean pesos. Data: Mineduc (2018).