This is a post-peer-review, pre-copyedit version of an article published in *Social Indicators Research*. The final authenticated version is available online at: https://doi.org/10.1007/s11205-021-02728-1

The role of catchment areas on school segregation by economic, social and cultural characteristics

Claudia Prieto-Latorre[1,2]

Oscar D. Marcenaro-Gutierrez[3]

Luis Alejandro Lopez-Agudo[4]

Abstract

This research analyses the socioeconomic and cultural segregation of students across school catchment areas using census data for the students in their second year of secondary education in Andalusia (the most populated region in Spain). The main methodology used is the Mutual Information Index, which satisfies all the desirable properties for measuring segregation. Concretely, we draw upon the additive decomposability property, which decomposes the segregation of students across schools into the different levels in which schools can be grouped, that is, catchment areas and, within catchment areas, by source of funding (public and semiprivate schools). We found that school segregation is greater than catchment areas' segregation. Additionally, statistically significant correlations are found between the level of segregation within the catchment areas and factors such as size of the catchment area, parental level of education and size of the municipality where the school is located.

Keywords: school segregation; socioeconomic and cultural level; catchment areas; mutual information index; public schools; semiprivate schools.

JEL codes: I24, R23.

The authors acknowledge the training received from the University of Malaga PhD Program in Economics and Business [Programa de Doctorado en Economía y Empresa de la Universidad de Malaga]. This work has been partly supported by FEDER funding (under Research Project UMA18FEDERJA024); Fundación Pública Andaluza Centro de Estudios Andaluces (under Research Project PRY085/19); the Andalusian Regional Government (SEJ-645) and Ministerio de Educación, Cultura y Deporte (scholarship FPU17/00432).

[1] Departamento de Economía Aplicada (Estadística y Econometría), Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, C/ Ejido, 6, 29013, Málaga, España. Tel.: +34 95 213 7003; Fax: +34 95 213 7262. Email: claudiaprieto@uma.es. ORCID: 0000-0002-6510-3057

[2] University of Malaga PhD Program in Economics and Business

[3] Departamento de Economía Aplicada (Estadística y Econometría), Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, C/ Ejido, 6, 29013, Málaga, España. Tel.: +34 95 213 7003; Fax: +34 95 213 7262. E-mail: odmarcenaro@uma.es. ORCID: 0000-0003-0939-5064

[4] Departamento de Economía Aplicada (Estadística y Econometría), Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, C/ Ejido, 6, 29013, Málaga, España. Tel.: +34 95 213 7003; Fax: +34 95 213 7262. Email: lopezagudo@uma.es. ORCID: 0000-0002-0906-3206

1. Introduction

The first goal of an education system is to ensure the access to education of youngsters and to provide inclusive and equitable education. In Spain, access to education is guaranteed for compulsory schooling (which lasts ten years), and this accessibility also enhances access to higher education (Consejo Escolar del Estado, 2018). However, the equity of the education system is still a questionable issue. The concept of equity in education has been widely addressed in the literature and all the approaches agree that an equitable education system promotes equal opportunities for students in the education process in terms of inputs, outputs and learning process (UNESCO, 2018; González-Betancor & Marcenaro-Gutiérrez, 2018). In this sense, many factors might perpetuate inequality in education (gender, ethnicity, religion, socioeconomic and cultural status, political economy, societal values and norms, etc.) which, together with their interactions, ultimately explain the existing educational disparities (Jacob & Holsinger, 2008).

Specifically, in this research we address the inequity in education arising as a consequence of student sorting by their economic, social and cultural status (hereafter ESCS). It is well established that ESCS has a strong influence on their academic performance (Sirin, 2005; Martins & Veiga, 2010; Cordero *et al.*, 2011); additionally, it has also been found that peer characteristics make a meaningful difference in students' academic achievement (Lavy *et al*. 2012; Gibbons & Telhaj, 2016). In this context, school diversity may serve as a mean to compensate the influence of socioeconomic and cultural characteristics on academic performance (Givord, 2019a). In fact, results from PISA[5] 2015 show that, in countries with a high level of school segregation, the association of ESCS with reading performance is greater and socioeconomic and cultural segregation have negative consequences; this happens especially in the case of the poorest students, as their initial disadvantage in terms of ESCS is not compensated by the education system (Krüger, 2018). In this respect, previous studies have found that exposure to a socially disadvantaged environment is negatively associated to students' likelihood of graduating from high school (Harding, 2003; Goux & Maurin, 2007; Wodtke *et al.*, 2011). Besides that, the influence of school socioeconomic and cultural composition remains important even after leaving high school, and long term consequences are considered to potentially have more influence on students' lives than short term ones (Palardy, 2015). In the short term, low income students who attend segregated schools (i.e. schools with a high concentration of low income students) have lower academic performance and more limited social interactions and, in the long run, attending these segregated school reduces the probability of going to university which, in theory, reduces students' potential for social mobility (Palardy, 2013; Klugman & Lee, 2019). Palardy (2013) found that attitudes, motivation and social norms may be the underlying drivers of these differences.

The present study seeks to identify the contribution of catchment areas and school funding to student's segregation across schools by socioeconomic and cultural status, in Andalusia (the most populated Spanish region). The first factor that may be driving school segregation are catchment areas, which are administrative divisions of the territory that are used to regulate school assignment. The proximity to the school is very important in the school admission process (the location of home or parental working place within the catchment area of a school increases the possibilities of having access to that school). In this way, the composition of catchment areas reflects the

[5] Programme for International Student Assessment.

socioeconomic and cultural structure of neighbourhoods and, thus, when analysing segregation across catchment areas, we tackle geographical segregation. The second source of school segregation is the presence of public and semiprivate schools. Public schools are fully financed by the state and have public ownership, while semiprivate schools are privately owned but receive most of their funding from the state. Although both types of school have the same admission policy, previous studies noted their different ESCS composition (Fernández-Enguita, 2008; Murillo & Martínez-Garrido, 2018).

In particular, unevenness is the dimension of segregation that is going to be explored (Massey, 2012); in the context of this study, unevenness measures the degree to which socioeconomic and cultural composition within social units (for instance, schools or catchment areas) departs from the socioeconomic and cultural composition of the entire population. When considering schools (or, alternatively, neighbourhoods or catchment areas) as social units, school segregation will be assessed (alternatively, neighbourhoods or catchment area segregation). More generally, students' segregation refers to the uneven distribution of students across social units. The larger the difference between the socioeconomic and cultural composition of social units and that of the population, the greater the level of students' segregation.

In order to carry out this research we make use of a census dataset which contains information about students attending secondary education schools in the largest region of Spain (Andalusia). It contains information related to students' performance, school and family characteristics, and the catchment area where the school is located. This research work is novel in at least three ways: first, information on catchment areas has never been employed to analyse this issue in the Spanish context. Second, the approach to analyse high school segregation at different levels has not been explored before in Spain. In particular, our approach considers that schools are nested within catchment areas. In order to undertake this analysis, we use the Mutual Information Index (Frankel & Volij, 2011). The advantages of this index have been documented and its decomposability properties clearly differentiate it from other measures of segregation (Mora & Ruiz-Castillo, 2011). Specifically, we use decomposability across units to simultaneously compute students' segregation across schools and catchment areas. Third, differences between public and semiprivate schools within catchment areas have not been explored before.

Consequently, this study sheds light about a new source of school segregation in Spain, i.e. catchment areas. So far, the unique source of school segregation explored in Spain is the coexistence of public and semiprivate schools, which only explained 25% of school segregation. This is due to the lack of data about catchment areas. We only found two recent studies which use information related to catchment areas in Spain, but with lower scope than this research (they analyse cities instead of regions); for the city of Madrid, Gortázar *et al*. (2019) study school segregation by parental level of education[6] and, for the city of Barcelona, Bonal *et al*. (2019) show the segregation of foreign students.

The paper is structured as follows. First, we present a review of the relevant literature related to segregation, with particular attention to the Spanish context. Second, the institutional background of the school admission process in Andalusia is introduced. We then describe the main features of the dataset and the methodology used. Lastly, we

[6] However, these authors indicated that parental level of education was not available at individual level, so they imputed the average level of the census block.

report the results and draw the main discussion and conclusions, including education policy implications.

2. Literature review

Literature has indicated that students' academic progression is not only influenced by the peer group composition at school, but also by the ESCS of the neighbourhood where they live. In this section, we review the literature related to both peer group effects and the socioeconomic and cultural characteristics of the neighbourhood. In addition, we refer to previous evidence on segregation in the Spanish context.

2.1. Peer group effect

Many research studies have found that peer group composition has an influence on students' academic attainment (Hanushek *et al*., 2003; Eisenkopf, 2010; Epple & Romano, 2011; Lavy *et al*. 2012; among others). In general, classmates with a high performance have been found to benefit the other students, while a high proportion of low achievers negatively influenced the academic attainment of their peers. This association is heterogeneous; for example, following Lavy *et al*. (2012), only "very bad" students (in the bottom 5% of the scores' distribution) present a negative influence from their classmates, while the peer effect on top performing students is less noticeable (Hanushek *et al*., 2003). The channels by which peer effects operate are multiple, e.g. it is plausible that students instruct each other or that peers influence one another via classroom disruption and classroom environment (Figlio, 2007). According to Vardardottir (2013, p. 108), it could simply be that "if students are sorted into classes based on their ability, it might allow teachers to match instructions more closely to students' needs because of more homogenous groups", though there is mixed empirical evidence to support this proposition.

Teachers' and learners' interaction also relate to students' education process and is conditioned by the social composition of the class (Harris & Williams, 2012). In more affluent schools, teachers may accept more open questions and spend more time answering them, which positively associates with students' academic performance. A high proportion of repeaters in the class – a condition which typically correlates with low ESCS (González-Betancor & López-Puig, 2016) – may also change teachers' behaviour, focusing more on memorising and less in developing students' skills (Lavy *et al*., 2011).

School peers' socioeconomic and cultural background also conditions students' academic attainment. A relatively high ESCS position in the school is positively associated with students' academic performance (commonly known as big-fish-little-pond effect). For the Spanish region of Andalusia, Jerrim *et al*. (2020) found that students' relative ESCS in primary education is a key determinant in their progression to secondary education, without any influence of the type of school (public and semiprivate schools).

2.2. Neighbourhood segregation

The reduction of income inequality levels in cities is not significantly associated with city inclusion, because the first variable tends to capture the dislocation of low income residents (Stacy *et al*., 2019). At the neighbourhood level, the concentration of low income families in distressed neighbourhoods has the same negative influence on their well-being as segregation has on this variable at the school level (Ludwig *et al*.,

2012). In fact, living in distressed neighbourhoods negatively conditions well-being rather than the socioeconomic and cultural level itself.

Individuals' residential mobility is constrained by the market mechanism of house prices (Brasington *et al*., 2015), but neighbourhood segregation also results from individuals' decisions, because people prefer living surrounded by others with an ESCS position similar to theirs (Musterd *et al*., 2016). Another determinant factor of students' performance is neighbourhood stability (Gibbons *et al*., 2017); students living in neighbourhoods with a high turnover seem to have a lower academic progression. Recently, the level of socioeconomic and cultural segregation in European cities has been measured on a relatively consistent basis: Musterd *et al*. (2017) conclude that, from 2000 to 2010, inequality in major European cities has increased and, if this trend continues, European cities will soon have similar levels of segregation as those of American cities. This differential level of segregation in European cities is due to the globalisation process, the welfare regime and the housing system. This upward trend in city segregation may result in an increase in school segregation (Denton, 1995; Croxford & Paterson, 2006) and a higher academic achievement gap, exacerbating the role of socioeconomic and cultural status in the learning process (Burger, 2019).

2.3. Segregation in Spain

The socioeconomic and cultural segregation across schools in Spain is not very high in secondary education compared to other OECD countries (Gutiérrez *et al*., 2019). Social groups are differently influenced; well-off students are more segregated than those with a low ESCS (Murillo *et al*., 2018; Prieto-Latorre *et al*., 2020). However, compared to Europe, Spain has one of the most segregated secondary education systems (Murillo & Martínez-Garrido, 2018). Spain is the third country with the highest level of segregation of high income students and the sixth in terms of low income students. The level of segregation in Spain varies across regions (Mancebón-Torrubia & Pérez-Ximénez de Embún, 2010). Gorard's segregation index and the isolation index have both shown that Andalusia has a medium level of segregation, similar to other Spanish regions such as Murcia and Valencia (Murillo & Martínez-Garrido, 2018).

Previous literature also notes the existence of a heterogeneous student distribution among schools according to their funding characteristics, leading to social segregation between and within school types. In other words, there seems to be a selection bias in school choice, conditioned by students' ESCS, as found by Fernández-Enguita (2008), Mancebón-Torrubia and Pérez-Ximénez de Embún (2009), Murillo *et al*. (2018) and Prieto-Latorre *et al*. (2020). In particular, these authors found that high ESCS students are concentrated in private and semiprivate schools. In this matter, private schools are not part of the debate because they do not receive public funding and are managed by market mechanisms. However, even though public and semiprivate schools receive public funding, they have a significant heterogeneous distribution of students. In line with these findings, research studies for Spain have identified differences in the level of segregation between public and private secondary education schools. Using data on PISA 2015 for Spain, Murillo *et al*. (2018) applied the Hutchen Index and found that differences between both types of schools entailed 22% of the level of segregation for poor students and 32% of the segregation for rich students[7]. Additionally, the segregation within each type of school has been analysed by Mancebón-Torrubia and Pérez-Ximénez de Embún (2009) using the Dissimilarity Index. These authors found that the segregation in semiprivate schools was greater than

[7] In their analysis they grouped semiprivate and private schools.

the segregation in public schools. This analysis considered schools as independent units, conversely to our study, in which we also take into account nearby schools through information on their catchment areas. In addition, these authors used the level of parental education as a proxy for ESCS, which constitutes a less precise measure of the socioeconomic and cultural profile of the students than those used in our analysis, in which parental occupation, the number of books at home and home educational resources are also considered through an index of economic, social and cultural status.

3. School admission process in Andalusia

In Spain, the national legislation establishes the basic regulation for students' access to the education system, the organisation and regulation of which is handled by Autonomous Communities (regions). The education legislation establishes that families can choose their children's school; however, most schools experience over-demand. Because of that, among the criteria provided in the Spanish law, regional legislation assigns a great weight to proximity to school in order to regulate student admissions.

Specifically, in Andalusia a point-based system determines students' assignment to school (as regulated in Law 53/2007 of 20th February 2007[8]), which is described in Table A1 (Appendix). Concretely, the proximity of the home or parental working place to the school stands as the most important criterion to have access to school (equivalent to 10 points if either one is located in the catchment area and 6 points in a boundary area). Students also have enrolment priority when they have siblings attending that school (6 points), a low family income (ranging from 2 to 0.5 points, depending on the income threshold they belong to), a "large family" status (families with three or more children) or single parent family (2 points), any of their parents working at the school (1 point), or when the student, parents or siblings present some kind of disability (2 to 0.5 points).

Delimitations of school catchment and boundary areas are based on the population density and educational facilities in the area. Each address (home or parental working place) is assigned to the catchment area of a single school, while boundary areas are more flexible. This school admission system does not discern between place of residence and parental working location, giving the same importance to both aspects. Nevertheless, a minor subsequent modification underlined the use of the place of residence, giving the latter greater weight than parental working place in the student school admission process (Law 40/2011, of 22nd February 2011, granted 4 more points to home proximity than to working place proximity); besides that, the admission criteria has remained unchanged in Andalusia.

The admission process concerns schools which receive public funding, that is, public and semiprivate schools. Public schools are fully publicly funded, whereas semiprivate schools receive around 70%[9] of their funds from State resources. In the academic year 2011-12 in Andalusia, 75.5% of secondary education students attended public schools and 22.2% attended semiprivate schools (Statistics of the Regional Government of Andalusia, 2012). The possibility of choosing between a public or

[8] Our dataset contains information related to 2nd year secondary education students in the academic year 2011/12, who accessed secondary education in the academic year 2010/11, the reason why we focus on this law, which was applicable at that moment.

[9] The Survey of Financing and Expenditure of Private Education (*Encuesta de Financiación y Gastos de la Enseñanza Privada*) is conducted every 5 years. In 2009-10, semiprivate schools were publicly funded in 75%, while they were in 69.2% in 2014-15.

semiprivate school is not covered by law; thus, admission to both types of schools is conditioned by the same geographical criteria mentioned above.

The limited autonomy that Spanish schools have in this regard, due to the centralised nature of regional education, would seem to imply that school choice *per se* is less likely to be a major source of segregation in the Spanish case (Pont *et al*., 2014). Home address to assign schools is a widely used policy across different education systems, but in Spain this criterion further accounts for the admission process compared to other OECD countries (Givord, 2019b). Therefore, the main factor that explains school segregation in Spain is the location of the student's home, along with the location of the parental working place. Thus, the emphasis placed on proximity to schools shows that school composition is highly associated to neighbourhood composition.

4. Data

In Andalusia, all students took two external standardised tests during compulsory education, one of them in 4th year of primary education and the other one in the 2nd year of secondary education. These exams did not determine students' progression; however, the information was used by educational authorities at an aggregated level for education policy purposes. These assessments were part of an annual evaluation carried out by the Andalusian Agency of Educational Assessment and evaluated students' competences in the core subjects of reading and mathematics. Additionally, background questionnaires gathered information about students' personal (answered by the student) and home characteristics (answered by parents), teachers (answered by the teacher) and school characteristics (answered by the school head teacher). In particular, we employ this census data for second year secondary education students in the academic year 2011-12 in Andalusia.

The Andalusian Agency of Educational Assessment constructed an economic, social and cultural status (ESCS) index – following the guidelines of the National Education Assessment for the second year of secondary education in 2010 (MECD, 2011) – which is the variable used as a measure of students' segregation. Concretely, parents' questionnaires, which have been found to be a good tool for capturing students' socioeconomic and cultural status (Schleicher *et al*., 2009), were used to create this index. The ESCS index comprises four components: the highest level of parental education, the highest level of parental occupation, the number of books at home and home educational resources. The first three variables are obtained directly from the parental questionnaire, while the last variable is built using the number of home possessions declared by parents[10]. The index is a continuous variable standardised to have mean 0 and standard deviation 1, using information from the population of Andalusian students. For the purpose of the analysis, the distribution of the ESCS is divided into 4 equal-size groups (i.e. quartiles), containing each one 25% of the total observations. The 1st quartile gathers the 25% of students from the lowest ESCS (i.e. the *poorest* students in socioeconomic and cultural terms), the 2nd quartile the next lowest 25% (up to the median), the 3rd quartile gathers the second highest 25% of students (25% immediately above the median) and the 4th quartile gathers the highest 25% of ESCS distribution (i.e. the *richest* students).

[10] These items are: the student has a suitable place to study at home; a desk; computer at home; Internet at home; resource books and education supporting books (dictionaries, encyclopaedia, etc.); books (novels, books of short stories, poems, comics, etc.); specialised magazines; daily press.

Other information in this dataset which allows quantifying students' segregation is school catchment areas, due to the importance they are assigned in the admission process. In our data, we can identify those schools which belong to the same catchment area and can thus analyse the characteristics of nearby schools. It is relevant to highlight that this information is available for the first time for a Spanish region.

In our initial dataset, in 2011-12 there were 90,849 students enrolled in the 2nd year of secondary education. These students attended 1,624 high schools organised into 486 catchment areas. For 17,474 students, the ESCS index was not available, so we could not include them in our analysis. Furthermore, the admission process in private schools is not regulated by the education legislation and less than 3% of Andalusian students attended private schools, so we did not include private schools (which are a total of 47) in our analysis.

In addition, this study is focused on catchment areas with more than 1 school (198 catchment areas are discarded). Although catchment areas with only 1 school represent 40% of the catchment areas of our sample, they only encompass 14% of the total amount of students (i.e. 10,065 students). These catchment areas have some characteristics which make them different from the rest, as presented in Table A2 (Appendix). Specifically, over half of these schools are located in rural areas (52% of schools are located in a municipality with fewer than 5,000 inhabitants and only 1% of schools belong to cities with more than 100,000 inhabitants). In these catchment areas there are only public schools (with the exception of one semiprivate school) and the socioeconomic and cultural level of students enrolled in catchment areas with 1 school is very low: on average, schools have 36% of disadvantaged students (1st quartile of the ESCS distribution) and only 12% from the 4th quartile. The variables which compose the ESCS index show a similar behaviour. In catchment areas with 1 school, 68% of fathers and mothers have not achieved a level of education higher than compulsory education; only 7% of parents have university studies while, in the other catchment areas, twice the number of fathers and mothers complete university studies. Given this situation, the core of our study does not include these one-school catchment areas.

Thus, the final sample comprises 61,557 students who attended 1,297 schools belonging to 279 catchment areas. Table A3 (Appendix) presents descriptive statistics for the catchment areas included in our analysis; they have an average of 4.6 schools, of which 80% are public schools and 20% semiprivate schools.

The differences between the socioeconomic and cultural composition of public and semiprivate schools are apparent and are presented in Table A4 (Appendix). The higher socioeconomic and cultural status of the semiprivate schools is perceived not only in the ESCS index, but also independently in each one of the variables comprising the ESCS. For example, in public schools, around 10% of parents studied for a university degree or a PhD, but this figure rises to 26% in semiprivate schools.

The relevance of addressing school segregation is based on the consequences that grouping students by socioeconomic and cultural level has on their academic performance. Descriptive statistics displayed in Tables A5 and A6 (Appendix) show that the ESCS composition of schools and catchment areas matters, i.e. students' academic performance *a priori* seems to be conditioned by the socioeconomic and cultural characteristics of their peers. The rows in these tables represent the standardised academic performance in reading and mathematics of students from the 1st and 4th quartile of the ESCS distribution, respectively, whilst the columns show the school's ESCS quartile (Table A5, Appendix) or the catchment area ESCS quartile (Table A6,

Appendix). In view of these tables, it seems that being surrounded by well-off students is positively related to academic performance for both high and low ESCS students. In terms of catchment areas, attending a school placed in a high or a low ESCS catchment area also entails a difference in terms of academic performance, but it is slightly lower than for school composition. It is important to take into account that richer students obtain better results in reading and mathematics than poorer students, regardless of the ESCS composition of the school. Therefore, disadvantaged students are those who seem to benefit the most from attending a school with a high ESCS level, since their improvement in academic performance places them near the average performance, while well-off students show top performance, regardless of the socioeconomic and cultural profile of their classmates. If students are sorted by ESCS level, the academic achievement gap which already exists by ESCS might be increased. Consequently, barriers which reduce students' mobility across catchment areas are a major obstacle on the academic progression of more disadvantaged students in terms of ESCS.

5. Methodology

5.1. The Mutual Information Index

The Mutual Information Index has been employed to assess the evenness of the socioeconomic and cultural distribution of students across schools. This index was first proposed by Theil and Finizza (1992). Besides satisfying basic desirable properties, decomposability properties clearly differentiate the Mutual Information Index from other measures of segregation, as we will see.

The Mutual Information Index (M Index from now on) is a multigroup index, i.e. it analyses the distribution of the population that is classified in two or more groups along social units. This is particularly desirable for the current research work, as the ESCS index is used to allocate students to four groups (i.e. quartiles of the ESCS distribution) and, after this, they are distributed amongst secondary schools. Following the notation used by Guinea-Martín *et al*. (2015), the total level of students' segregation across schools using the M Index is obtained as follows:

$$M = \sum_s p_s\left[E\left(P_q\right) - E\left(P_{q|s}\right)\right] \quad (1)$$

where:

$s = 1,2,3,4 \ldots, n$ is the number of schools.

$q = 1,2,3,4$ is the quartile of the ESCS distribution.

p_s is the proportion of students in school s.

P_q represents the distribution of students by quartiles of the ESCS distribution.

$P_{q|s}$ is the conditional distribution of students by quartiles of the ESCS in school s.

$E\left(P_q\right)$ is the entropy of the distribution of students by quartiles of the ESCS distribution.

$E\left(P_{q|s}\right)$ is the entropy of the distribution of students by quartiles of the ESCS in school s.

This M Index can measure the distribution of students grouped in two or more subgroups across schools. As our interest is to know the overall distribution by ESCS, but also the distribution of particular subgroups, we also use a binary version of this index. This version involves defining two groups instead of four. For instance, we will

compare the distribution of students who belong to the first quartile of the ESCS distribution (i.e. the poorest students) with the rest of subgroups. Following the previously used notation, the variable q, instead of taking 4 values, will take only 2, that is, 1st quartile vs the rest of ESCS subgroups (students from the 2nd, 3rd and 4th quartiles). The same procedure is applied to the rest of subgroups, i.e., the comparison of students from the 2nd quartile vs the rest of ESCS groups (students from the 1st, 3rd and 4th quartiles), and so on for the 3rd and 4th quartiles.

As it can be seen, the M index is based on the concept of entropy of a distribution. Entropy is defined as "the uncertainty of a single random variable" (Cover & Thomas, 1991, p. 6). For a discrete random variable x, being i its different values and p_i the probability that it takes the value i, the entropy $\left(E(x)\right)$ is:

$$E(x) = -\sum_i p_i \, log \, p_i \qquad (2)$$

Information and entropy go in opposite directions, i.e. a higher probability of taking the value i implies a lower entropy and *vice versa*. The distribution of students by quartiles is uniform (as the proportion of students in each ESCS quartile, by definition, is the same), and achieves the maximum entropy. The index compares the entropy of the groups of students with the conditional distribution of groups of students within the school. The school the student attends reduces the uncertainty of the distribution of the groups of students and is called mutual information (Cover & Thomas, 1991). The term $\left[E\left(P_q\right) - E\left(P_{q|s}\right)\right]$ is the reduction in uncertainty of the distribution of students by quartiles due to the school.

The lower bound of the index is 0, which means that there is a lack of segregation, so the proportion of students from each quartile is the same within each school, i.e. the distribution of students by ESCS and the distribution of students by ESCS conditioned by the school are the same $\left(E\left(P_q\right) = E\left(P_{q|s}\right)\right)$. The upper bound of the index is not fixed; it is obtained as the natural logarithm of the number of groups[11] (Guinea-Martin *et al*., 2015). In case students are grouped into 4 quartiles of ESCS, the maximum value of the M index is 1.386 (or the natural logarithm of 4) and, when students are grouped into two groups, the maximum value is 0.693 (the natural logarithm of 2).

The M index satisfies the following basic properties, which are desirable for a measure of segregation (see Frankel & Volij, 2011, for an axiomatic characterisation of the properties of the M Index):

(a) It is scale invariant, which means that the index does not vary when the number of students increases proportionally across groups within each school.

(b) It meets the principle of organisation equivalence, which allows to compare the segregation of two catchment areas with different number of schools and different number of students. This is important, because it ensures that variability in the index does not arise from differences in the number of schools or students within catchment areas.

(c) Size invariance property ensures that the combination of two schools with the same ESCS distribution gives the same level of segregation.

[11] The upper bound is the lower value of either the logarithm of the number of groups (4 or 2 groups) or the logarithm of the number of schools (1,297 schools).

(d) It accomplishes the school division property, which states that a school can be divided into sub-schools and the level of segregation will never be lower than that of the original school. In case the school is divided into sub-schools with the same share of students across groups, the level of segregation is not altered (insensitivity to proportional division); otherwise, the index will change.

(e) It follows the transfer principle, which describes the movement of a student from any group (any quartile of the ESCS distribution) to another school. In the multigroup approach we cannot foresee whether the value of the index increases or decreases after this shift, but so we can in the binary approach.

(f) It presents symmetry in groups and symmetry in units, which indicate that the labels of the ESCS quartile and the school name do not influence the index value.

(g) The index is transpose-invariant, that is, it treats the quartile the student belongs to and the school the student attends in a symmetrical way and, hence, quartiles and schools can be transposed (or interchanged).

(h) It follows the zero member property, which indicates that segregation remains unchanged if a school without students is added or removed.

(i) It is not composition invariant, which implies that if students from a particular ESCS quartile (for example, students from the 1st quartile, i.e. the "poorest" students) increase or decrease proportionally across schools, the M index will change. The desirability of this property is not clear (for a discussion see Gorard & Taylor, 2002; Allen & Vignoles, 2007). Although this property is not very relevant for our application, as we are using a relative measure of students' socioeconomic and cultural level in a cross section analysis, if we used a fixed boundary to classify students, such as a fixed income threshold in panel data (i.e. for more than one period), the implications of this property could be problematic. This is because a strong composition-invariant index would not reflect, for instance, the influence of a financial crisis which genuinely increased the number of poor students, if the number of students who switched to poor status within schools was proportional to the increase in the number of poor students in the population.

(j) It meets the property of strong group decomposition, which allows the analysis of the contribution of each group (i.e. each quartile) to the level of segregation.

In the following subsection we present another property of the Mutual Information Index which is related to its decomposition.

5.2. The Mutual Information Index in our setting: the strong decomposability across units property

In the first step of our research, we will take advantage of the decomposition properties that the M index satisfies. In particular, the possibility of decomposing the level of segregation across units is especially relevant, because we want to identify the different sources of students' segregation at three levels simultaneously. This property is called strong decomposability across units, and it enables us to obtain the contribution to segregation of all the levels simultaneously (Mora & Ruiz-Castillo, 2011). The region of Andalusia is divided into school catchment areas, so the first level of segregation is the school catchment area. If we consider all public schools within the catchment area as one social unit and the set of semiprivate schools as the other social unit, the differential distribution of students across both types of schools would be the second level of segregation. Lastly, we identify segregation between public schools within a

catchment area and between semiprivate schools within a catchment area, which constitutes the third segregation level.

According to the strong decomposability across units property, the M index (M) can be decomposed into two terms: a first component measures the differential distribution of students across catchment areas (between; $M_{between\ catchment\ areas}$) and a second term measures how much segregation arises from the distribution of students within schools of a catchment area (within; $M_{within\ catchment\ areas}$).

$$M = M_{between\ catchment\ areas} + M_{within\ catchment\ areas} \quad (3)$$

The within component, using the aggregation property that this index meets (which is analogous to the independence property according to Frankel & Volij, 2011) is obtained as the sum of the school segregation within each catchment area weighted by the proportion of students within the catchment area:

$$M_{within\ catchment\ areas} = \sum_{c=1}^{279} p_c\, M_c \quad (4)$$

where:

$c = 1, 2, \ldots 279$ is the number of catchment areas.

p_c is the proportion of students in the catchment area c.

M_c is the Mutual Information Index of the catchment area c.

Using the property of strong decomposability across units, the M index within the catchment area c (M_c) can be decomposed into two terms: differences between public schools and semiprivate schools $\left(M_{c,between\ public-semiprivate}\right)$ and differences within public schools and semiprivate schools of a catchment area $\left(M_{c,within\ public-semiprivate}\right)$:

$$M_c = M_{c,between\ public-semiprivate} + M_{c,within\ public-semiprivate} \quad (5)$$

Again, the latter term $\left(M_{c,within\ public-semiprivate}\right)$ can be divided into the sum of segregation within public $\left(M_{c,public}\right)$ and semiprivate schools $\left(M_{c,semiprivate}\right)$, weighted by the size of each subsystem ($p_{c,public}$ and $p_{c,semiprivate}$, respectively):

$$M_{c,within\ public-semiprivate} = p_{c,public}\, M_{c,public} + p_{c,semiprivate}\, M_{c,semiprivate} \quad (6)$$

where $p_{d,public}$ and $p_{d,semiprivate}$ are the proportion of students within the catchment area c who attend public and semiprivate schools, respectively. $M_{c,public}$ and $M_{c,semiprivate}$ are the Mutual Information Index of the catchment area c which measure the ESCS distribution between public schools and semiprivate schools, respectively.

Other segregation indexes satisfy the additive decomposability across units, such as the Atkinson Index and the Hutchens Index. This property is less strict than the strong decomposability across units property, because it assumes that weights are defined as a function of the proportion of students, while the strong decomposability assumes that weights are equal to these proportions[12]. This property is a unique feature of the M Index (Frankel & Volji, 2011) and will let us analyse school segregation across

[12] For example, in the Hutchens Index, weight is a function which depends on the proportion of students in each group within the social unit (only for two groups, because it is based in a binary distribution of the population) and the proportion of students in each group in the population, while weight is the proportion of students who belong to the social unit in the M index.

catchment areas for the particular case of secondary education students in Spain for the first time.

5.3. Empirical Approach

The first step of our empirical strategy is to decompose school segregation into different sources. Figure A1 (Appendix) sums up the analysis explained below. Socioeconomic and cultural segregation of students across schools (M in equation 3) can be decomposed into segregation between catchment areas ($M_{between\ catchment\ areas}$ in equation 3) and segregation within catchment areas ($M_{within\ catchment\ areas}$ in equation 3). On the one hand, segregation between catchment areas compares the socioeconomic and cultural composition of catchment areas, considering all schools within the same catchment area as a unit, regardless of their source of funding (i.e. public and semiprivate). On the other hand, differences within a catchment area arise because of two factors. Firstly, public and semiprivate schools within a catchment area have different socioeconomic and cultural composition ($M_{c,between\ public-semiprivate}$ in equation 5). To measure this first source of segregation, we consider public schools within a catchment area as one social unit and semiprivate schools as the other social unit. This means that, if a catchment area has only one type of school (for example, there are only public schools), there is no segregation at this level. However, in the case that a catchment area contains both types of schools, at this level we would measure different distributions between these two units. The second source of segregation would be the distribution of students among the same type of schools within a catchment area ($M_{c,within\ public-semiprivate}$ in equation 5), which can be decomposed into segregation between public schools and segregation between semiprivate schools, both respectively weighted by the proportion of students attending public and semiprivate schools ($p_{c,public}\ M_{c,public}$ and $p_{c,semiprivate}\ M_{c,semiprivate}$, respectively), as shown in equation 6.

The second step of the analysis is to associate the level of segregation within catchment areas to some characteristics of the students and the schools which comprise that catchment area. Some factors which may be driving segregation within a catchment area are the average ESCS levels of the catchment area, the size of the municipality where the catchment area is located and the number of schools within the catchment area, among others. Using the M index of socioeconomic and cultural segregation of each catchment area (M_c) – described in equation 5 – we associate the features of the catchment area with the ESCS composition of schools which belong to that catchment area, using correlation coefficients and also an Ordinary Least Squares estimation:

$$\rho = \frac{Cov(M_c, F_c)}{\sigma_{M_c} \cdot \sigma_{F_c}} \tag{7}$$

$$M_c = \alpha + \beta \cdot ESCS_c + u_c \tag{8}$$

where:

M_c is the segregation within the catchment area c.

F_c is the feature of catchment area c (average socioeconomic and cultural level of the catchment area, proportion of parents with university studies within the catchment area, the size of the municipality where the catchment area is located, number of schools within the catchment area and proportion of public and semiprivate schools within the catchment area).

Cov represents the covariance.

σ_{M_c} and σ_{F_c} are the standard deviations of M_c and F_c, respectively.

ρ is the Pearson's correlation coefficient.

$ESCS_c$ is the ESCS of the catchment area *c*.

u_c is the idiosyncratic error term.

α and β are regression coefficients.

6. Results

6.1. First step: Mutual information index decomposition

This section presents the main findings on the socioeconomic and cultural segregation of students across schools in Andalusia using the M Index. Table 1 reports the socioeconomic and cultural segregation of students at the different levels. The first column of the table computes students' segregation from the whole distribution of the ESCS index (the four quartiles simultaneously), while the rest of the columns consider a binary distribution of students in order to identify possible differential contributions from students of different socioeconomic and cultural groups. Focusing on students' segregation from the whole distribution of the ESCS index, the value of socioeconomic and cultural segregation across the four quartiles is 0.174. The fact that the upper bound of the M index is not fixed makes it difficult to easily appreciate the level of segregation. Taking into account that the maximum value of the index for segregation across quartiles is 1.386, we found that the level of segregation reaches 12.55% of its maximum value.

The value of school segregation (0.174) can be decomposed into segregation between catchment areas (0.082) and segregation within catchment areas (0.092). As it is not readily apparent how much each part contributes to the total level of school segregation, Table 2 shows the contribution to each level. The percentages represent the share of segregation attributed to the different sources. In general, for computing percentages, we use the total school segregation (i.e. using the first row of Table 1, "Total: School segregation"). If we look at the decomposition expressed in percentages (Table 2), we observe that 47% of the segregation (0.082 out of 0.174) is due to the distribution of students between catchment areas and 53% (0.092 out of 0.174) stems from the distribution within catchment areas. Next, Table 2 decomposes the segregation within catchment areas (i.e. between public and semiprivate schools and within public and semiprivate schools) as a share of the total segregation (which is calculated using the first row of Table 1) and as a share of the segregation within the catchment area (values indexed with superscript A in Table 2, which are calculated using the third row of Table 1). Surprisingly, within catchment areas, the social composition of public and semiprivate schools represents 26% of the segregation (0.024 out of 0.092), and most of the segregation within a catchment area (74%) is due to differences within public and semiprivate schools (0.068 out of 0.092). Lastly, public and semiprivate schools differently contribute to the segregation within schools with the same type of funding. In this case, the reference level is segregation within the same school funding (values indexed with superscript B, which are calculated using the fifth row of Table 1). Differences between public schools account for 60% of the segregation within the same type of school in a catchment area (0.041 out of 0.068), while the remaining 40% (0.027 out of 0.068) is due to differences between semiprivate schools[13]. In fact, the

[13] In absolute terms, differences between semiprivate schools within the same catchment area are greater than differences between public schools within the same catchment area. For the total level of segregation

distribution of students across public schools within a catchment area explains 24% of the total level of school segregation (0.041 out of 0.174).

Moreover, the low weight of the segregation between public and semiprivate schools within a catchment area (around 26% of the segregation between schools within catchment areas; Table 2, first column, with the superscript A) could be a consequence of the high proportion of catchment areas where there is only one type of school: in 52.7% (147 out of 279) of catchment areas there are only public schools or only semiprivate schools. Nevertheless, after computing the Mutual Information Index for the subsample of catchment areas with both public and semi-private schools[14], there is only a slight increase in this component – from 26% to 31%. Hence, this does not seem to be a reason which explains the low contribution to the level of segregation between public and semiprivate schools within a catchment area, so we sustain that differences between the same type of school within the same catchment area are the most important source of segregation within the catchment area.

Besides, we have computed the level of segregation of catchment areas with only one school which, as it was previously argued, were not included in the core of our study[15]. The total segregation between schools by quartiles for this subsample is 0.091, which is about half of the total level of segregation between schools of catchment areas with 2 or more schools (0.174, Table 1).

Table 1. Segregation of students by quartiles of the ESCS distribution. Absolute values

	Segregation among 1st, 2nd, 3rd, 4th quartiles	**1st quartile vs {2nd, 3rd, 4th quartiles}**	**2nd quartile vs {1st, 3rd, 4th quartiles}**	**3rd quartile vs {1st, 2nd, 4th quartiles}**	**4th quartile vs {1st, 2nd, 3rd quartiles}**
Total: School segregation	0.174	0.077	0.033	0.022	0.105
Between catchment areas	0.082	0.035	0.014	0.007	0.055
Within catchment areas	0.092	0.042	0.019	0.015	0.050
• **Between public and semiprivate schools**	0.024	0.012	0.004	0.002	0.015
• **Within public and semiprivate schools**	0.068	0.030	0.015	0.013	0.035
o **Between public schools**	0.041	0.020	0.008	0.008	0.019
o **Between semiprivate schools**	0.027	0.010	0.007	0.005	0.016

Notes: The segregation of public and semiprivate schools within a catchment area is weighted by the proportion of students attending public and semiprivate schools, respectively. The percentage of students attending public schools is 71.86% and the percentage of students attending semiprivate schools is 28.14%.
Source: Authors' own elaboration.

Table 2. Segregation of students by quartiles of the ESCS distribution. Percentage decomposition by column

(Table 1, column 1), the M Index for public schools in the same catchment area ($M_{c,public}$) is 0.057 and the M Index for semiprivate schools in the same catchment area ($M_{c,semiprivate}$) is 0.095. However, these figures are weighted by the proportion of students attending public and semiprivate schools, respectively (only 28% of students attend semiprivate schools, this is, $p_{c,semiprivate}$), and this fact explains the higher contribution of public schools to the whole level of segregation.

[14] The decomposition of the Mutual Information Index for the subsample of catchment areas in which there are both types of schools (public and semiprivate schools) is available in the online supplemental material in Tables S1 and S2 (replicating Tables 1 and 2, respectively).

[15] The results of the Mutual Information Index for catchment areas with 1 school are available in the online supplemental material in Table S3 (replicating the first row of Table 1).

	Segregation among 1st, 2nd, 3rd, 4th quartiles	1st quartile vs {2nd, 3rd, 4th quartiles}	2nd quartile vs {1st, 3rd, 4th quartiles}	3rd quartile vs {1st, 2nd, 4th quartiles}	4th quartile vs {1st, 2nd, 3rd quartiles}
Total: School segregation	100.00%	100.00%	100.00%	100.00%	100.00%
Between catchment areas	47.13%	45.45%	42.42%	31.82%	52.38%
Within catchment areas	52.87%	54.55%	57.58%	68.18%	47.62%
• **Between public and semiprivate schools**	13.79% [A]26.09%	15.58% [A]28.57%	12.12% [A]21.05%	9.09% [A]13.33%	14.29% [A]30.00%
• **Within public and semiprivate schools**	39.08% [A]73.91%	38.96% [A]71.43%	45.45% [A]78.95%	59.09% [A]86.67%	33.33% [A]70.00%
o **Between public schools**	23.56% [B]60.29%	25.97% [B]66.67%	24.24% [B]53.33%	36.36% [B]61.54%	18.10% [B]54.29%
o **Between semiprivate schools**	15.52% [B]39.71%	12.99% [B]33.33%	21.21% [B]46.67%	22.73% [B]38.46%	15.24% [B]45.71%

Note: In general, the reference level to compute percentages is the total segregation between schools (calculated using the first row of Table 1, "Total: School segregation"). There are some exceptions: [A] The reference level is segregation within catchment areas for each column (third row of Table 1). [B] The reference level is segregation within the same type of school within a catchment area for each column (fifth row of Table 1).
Source: Authors' own elaboration.

The segregation by subgroup is quite different. Students from both ends of the ESCS distribution (with values of 0.077 and 0.105 for the 1st and 4th quartiles, respectively) are more segregated than those around the median of the ESCS (with values of 0.033 and 0.022 for the 2nd and 3rd quartiles, respectively). Besides that, students from the 4th quartile (the richest) are more segregated than students from the 1st quartile (the poorest). Taking into account that the maximum value of the binary index is 0.693, the segregation of the richest students is 15.15% of their maximum value, whilst the segregation of the poorest students is 11.11% of the maximum value.

The decomposition of segregation across levels is somewhat different by quartile (Table 2). In particular, we found differences in segregation between the richest students (4th quartile) and the other socioeconomic and cultural groups. The segregation between catchment areas is the most important factor accounting for segregation of the richest students (52% of segregation) while, for the other ESCS thresholds, differences within catchment areas are more important (over 50%). To the extent that catchment areas reflect neighbourhoods' composition, this highlights the more uneven distribution of the richest students across neighbourhoods than the rest of socioeconomic and cultural groups. Besides that, results show a different contribution of public and semiprivate schools to the level of segregation by socioeconomic and cultural group. Concretely, the main sources of segregation within catchment areas for 1st, 2nd and 3rd quartiles are differences between public schools (25.97%, 24.24% and 36.36%, respectively). In the case of students from the 4th quartile, the distribution of students between public and semiprivate schools has more or less equal weight (18.10% and 15.24%, respectively). A possible reason of this result is the concentration of students from the 4th quartile in catchment areas with a high proportion of semiprivate schools.

Another noteworthy fact is that relevant differences are identified between students over the median of the ESCS distribution, that is, students from the 3rd and 4th quartiles. First of all, students from the 3rd quartile have the lowest level of segregation (0.022, Table 1), conversely to students from the 4th quartile, who are the most segregated (0.105, Table 1). The size of the segregation between catchment areas is the lowest for students from the 3rd quartile (31.82%, Table 2), but the highest for students in the 4th quartile (52.38%, Table 2). If we study the level of segregation within catchment areas and, particularly inside it, we analyse the share of the segregation

between public and semiprivate schools (which presents the superscript A in Table 2), the lowest contribution comes from students from the 3rd quartile (13.33%, Table 2), whereas the highest contribution comes from students from the 4th quartile (30%, Table 2).

6.2. Second step: Association of the level of segregation within catchment areas with student and school characteristics

Differences within catchment areas are an important source of socioeconomic and cultural segregation; in fact, they entail more than half of the school segregation (52.87%, Table 2). That is why we delve into the factors which may condition the differential ESCS composition within the same catchment area. Thus, we compute the correlation between the M index within catchment areas (M_c) and some features of the neighbourhood (F_c) in Table 3. The first variable is the socioeconomic and cultural level of the catchment area, which is significantly and positively correlated to the M index within the catchment area (the pairwise correlation coefficient is 0.38). Figure 1 shows that, although there is some dispersion, there exists a positive association between both variables. Additionally, we analyse the correlation of the M index with the parents' level of education, another proxy of students' family background. The variable is defined as the proportion of parents with university studies within a catchment area, and we find that it is also significantly associated to the M index (0.40).

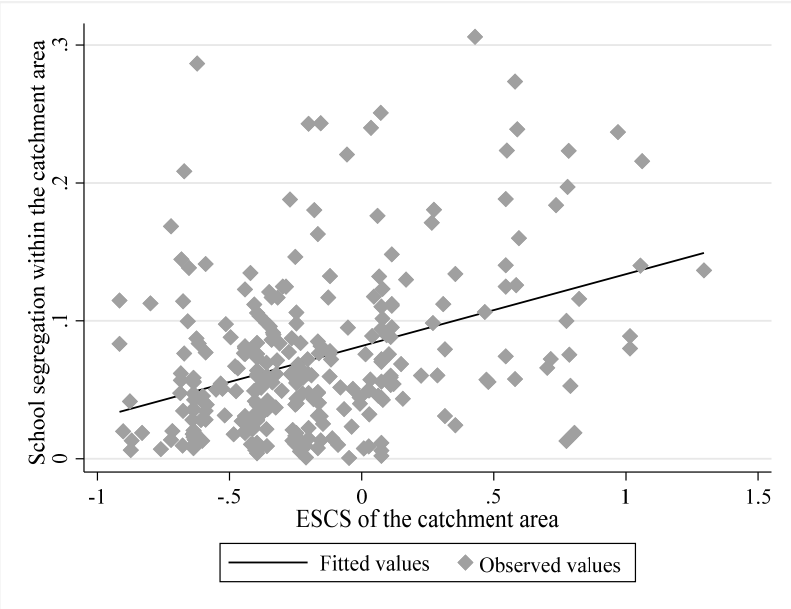
The size of the municipality where the catchment area is located is also associated to the segregation within the catchment areas. Municipalities with less than 50,000 inhabitants are negatively associated to segregation, whereas municipalities with more than 100,000 inhabitants have a positive correlation. Another variable that can be linked to segregation within catchment areas is the number of schools within the catchment area. We found that the higher the number of schools within a catchment area, the higher the level of segregation. Finally, the funding of schools comprising the catchment area is another source of segregation. A higher proportion of semiprivate schools positively correlates to the school segregation within the catchment area. Therefore, this high correlation of all these factors with the M index makes it difficult to identify the main factor which influences segregation.

Table 3. Pairwise correlation by catchment area of the M index of segregation within catchment areas and factors promoting segregation within catchment areas

	M Index within the catchment area
Socioeconomic and cultural level of the catchment area	0.377***
Proportion of parents with university studies within the catchment area	0.402***
Catchment area located in a municipality with less than 5,000 inhabitants	-0.097
Catchment area located in a municipality with 5,000 to 10,000 inhabitants	-0.189***
Catchment area located in a municipality with 10,000 to 50,000 inhabitants	-0.289***
Catchment area located in a municipality with 50,000 to 100,000 inhabitants	0.085
Catchment area located in a municipality with more than 100,000 inhabitants	0.538***
Number of schools within the catchment area	0.581***
Proportion of public schools within the catchment area	-0.483***
Proportion of semiprivate schools within the catchment area	0.483***

Note: "M Index within the catchment area" is the Mutual Information Index within the catchment area (M_c) by quartiles of the ESCS distribution (segregation among 1st, 2nd, 3rd and 4th).
Coefficient: ***Significant at 1%.
Source: Authors' own elaboration.

Figure 1. Scatter plot with regression line of the M index of segregation within catchment areas and the socioeconomic and cultural level of the catchment areas



Note: The *y*-axis is the Mutual Information Index of school segregation within a catchment area across quartiles of the ESCS distribution (segregation among 1st, 2nd, 3rd and 4th quartiles) and the *x*-axis is the ESCS of the catchment area. Fitted values are obtained from the regression model: $M_c = \alpha + \beta \cdot ESCS_c + u_c$, where $\alpha = 0.082$ (p-value=0.00), $\beta = 0.052$ (p-value=0.00), M_c is the Mutual Information Index within the catchment area c, $ESCS_c$ is the ESCS of the catchment area c and u_c is the idiosyncratic error term.
Source: Authors' own elaboration.

7. Discussion

This research study helps to disentangle the sources of socioeconomic and cultural segregation in Spain. Previous studies had identified one source of segregation: the coexistence of public and semiprivate schools (Gutiérrez *et al*., 2019; Murillo *et al*., 2018; Prieto-Latorre *et al.,* 2020). Existing literature had attributed around 25% of school segregation to differences between public and semiprivate schools, while the remaining 75% of segregation has been related to differences within public and semiprivate schools and, hence, other factors may explain it. This research addresses this issue for the first time and found that 47% of school segregation is due to catchment areas' segregation, 14% due to the coexistence of public and semiprivate schools within catchment areas and 39% are differences between public schools and semiprivate schools within the catchment area. Therefore, public and semiprivate schools do not seem as important as it was previously found (their relative importance diminished from 25% to 14%) and school catchment areas emerged as the main driver of school segregation. Probably, the declining share of segregation between public and semiprivate schools may be due to the location of semiprivate schools in catchment areas with a high ESCS and, hence, the segregation of the catchment areas picks up part of the segregation previously attached to differences between public and semiprivate schools.

Additionally, results also showed that the level of segregation differs between socioeconomic and cultural groups. In particular, it is worth to highlight that the richest students (4th quartile of the ESCS) are more segregated than the poorest (1st quartile). This finding is in line with previous literature: in Spain, the richest students have been found to be systematically more segregated than the poorest students (Gutierrez *et al*., 2019; Murillo *et al*., 2018). The high concentration of rich students is also specially

marked in Latin American countries and other European countries such as Hungary and Romania (Givord, 2019a). In fact, the level of segregation of the richest students in Spain is above the average of the OECD countries, while the segregation of the poorest students is around the average (Murillo & Martínez-Garrido, 2018).

Besides, this study contributes to the current debate on school admission policies. Although the trend is to increase parental freedom in respect to school choice, the social implications of this change are not so clear (Jeynes, 2000). Advocates of school choice consider that greater freedom for school choice is positively associated to school productivity (Hoxby, 2000) and that it also positively relates to graduation rates, because students can find a school which better meets their needs (Cullen *et al*., 2015). Conversely, those who are opposed to freedom of choice for families hold that it is positively associated to school segregation (Söderström & Uusitalo, 2010) and that it "reduces the unique potential of schools as social cohesion builders" (Musset, 2012, p. 4). Other studies have found that school choice policies do not have a significant influence on academic performance and segregation (Hsieh & Urquiola, 2006; Lindbom, 2010).

Lotteries as a method of distributing school admissions provide equal access to schools since it is a way of ensuring an "even social mix" (Simon *et al*., 2007). They have been applied to allocate students within school catchment areas (Musset, 2012), for the distribution of vouchers which partially covered school fees (Angrist *et al*., 2002; Shakeel *et al*., 2016), but they are mostly applied to solve the problem of oversubscribed schools (Stasz & Von Stolk, 2007; Deming *et al*., 2014). For instance, in some cities of England, lotteries replaced the criterion of proximity to school for oversubscribed places (Allen *et al*., 2013). Evidence about the implications of lotteries on ESCS equity is limited because most studies focus on the impact on academic outcomes (Stasz & Von Stolk, 2007). The study of Glazerman and Dotter (2017) addresses this issue; it simulates the impact of lotteries on students sorting and concludes that they would reduce socioeconomic, cultural and racial segregation. Because of that, we consider that random allocation of students within catchment areas could reduce the segregation within catchment area. In the same vein, voucher programs which subsidise fees of semiprivate schools could reduce the segregation arising from the coexistence of public and semiprivate schools.

8. Conclusions

The socioeconomic and cultural segregation of students influences the equity of an education system, since it reinforces the relative advantage or disadvantage of students that already exists due to socioeconomic and cultural background. In this paper we explore the segregation of students across schools using administrative data for the largest region of Spain (Andalusia) using the Mutual Information Index which – to the best of our knowledge – is the most suitable index of segregation to address this issue. We have found that school segregation is higher than catchment area segregation. The distribution of students across catchment areas represents 47% of the level of school segregation, while the remaining 53% stems from differences within catchment areas, which in turn depends on the existence of public and semiprivate schools and differences within each type of school. Besides, school segregation is different between socioeconomic and cultural groups. In particular, the main source of segregation rests on the basis of catchment area segregation for students in the 4th quartile of the ESCS distribution and, for the rest of ESCS quartiles, the segregation within a catchment area explains a higher proportion of school segregation. Focusing on the segregation within a catchment area, we have identified potential factors which could correlate to this

segregation: school segregation seems higher in catchment areas with a higher ESCS level, those where there are more semiprivate schools and catchment areas located in big cities.

There are many factors behind these findings. First, the importance of home proximity in school admission notes that the segregation found (high level of segregation between catchment areas) is a clear consequence of catchment area segregation. At the same time, the school admission policy stresses the barriers between neighbourhoods. However, removing catchment areas does not seem to be an alternative. School choice opportunity should be carefully considered, since previous studies have found that it intensifies inequities (Musset, 2012). Thus, modifying catchment areas may balance the current policy and school choice. In this sense, a redesign of catchment areas, taking into account the socioeconomic and cultural level of the families or the expansion of catchment areas, is an option. However, the implementation of this initiative should be coupled with parental information to avoid having high ESCS families take advantage of this measure. Furthermore, parental information could explain the high segregation within catchment areas. In fact, the differential access to information that parents from different ESCS have about school characteristics is one of the reasons which justifies that school choice opportunities increase segregation (Hasting & Weinstein, 2008).

Considering that catchment areas remain as they are, the presence of public and semiprivate schools within a catchment area would be another source of segregation. Even though this presence only represents 14% of the whole level of segregation, its relative importance increases up to 26% when we consider the segregation within a catchment area. Both types of school share the same legal basis; however, extra-costs associated to extracurricular activities in semiprivate schools would be driving the sorting of students by socioeconomic and cultural background. A simple way to remove this source of segregation would be to unify students' fees for both school types. However, the funding of semiprivate schools largely relies on private funding (30% of their funds are not public), so this would not be possible. More difficult to implement, but still feasible, would be to condition students' fees to family ESCS in semiprivate schools (Prieto-Latorre *et al*., 2020).

Schools with a high concentration of low ESCS students face an additional problem based on attracting highly qualified and experienced teachers (Martínez & Ferrer, 2018). This is another factor which increases inequity in education. In addition, these schools have a high mobility of teachers. As the working environment of teachers is more challenging, incentives could be used to attract and retain teachers in these schools, not only in economic terms, but also by reducing the teaching burden; both could help deal with this problem of segregated schools and to create a stable teacher staff.

All in all, more economic and cultural resources should be allocated to schools and neighbourhoods with a high proportion of low ESCS students, which are those that suffer the negative consequences of segregation.

Although we have employed an index with desirable properties and census data, this research work is not free of limitations. First, our results on the determinants of segregation are correlational and, thus, should be interpreted with caution. Second, although our results may have high internal validity for the Spanish region of Andalusia, they do not have much external validity; hence, this issue should be explored for other regions.

References

Allen, R., Burgess, S., & McKenna, L. (2013). The short-run impact of using lotteries for school admissions: early results from Brighton and Hove's reforms. *Transactions of the Institute of British Geographers*, *38*(1), 149-166. doi: 10.1111/j.1475-5661.2012.00511.x

Allen, R., and Vignoles, A. (2007). What should an index of school segregation measure?. *Oxford Review of Education*, *33*(5), 643-668. doi: 10.1080/03054980701366306

Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American economic review*, *92*(5), 1535-1558. doi: 10.1257/000282802762024629

Bonal, X., Zancajo, A., & Scandurra, R. (2019). Residential segregation and school segregation of foreign students in Barcelona. *Urban Studies*, *56*(15), 3251-3273. doi: 10.1177/0042098019863662

Brasington, D. M., Hite, D., & Jauregui, A. (2015). House price impacts of racial, income, education, and age neighborhood segregation. *Journal of Regional Science*, *55*(3), 442–467. doi: 10.1111/jors.12173

Burger, K. (2019). The socio-spatial dimension of educational inequality: A comparative European analysis. *Studies in Educational Evaluation, 62*, 171–186. doi: 10.1016/j.stueduc.2019.03.009

Consejo Escolar del Estado (2018). *Informe 2018 sobre el estado del sistema educativo*. Madrid: Ministerio de Educación y Formación Profesional.

Cordero, J. M., Crespo, E., Pedraja, F., & Santín, D. (2011). Exploring educational efficiency divergences across Spanish regions in PISA 2006. *Revista de economía aplicada*, *19*(57), 117–146.

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory*. New York: John Wiley & Sons.

Croxford, L., & Paterson, L. (2006). Trends in social class segregation between schools in England, Wales and Scotland since 1984. *Research Papers in Education*, *21*(4), 381–406. doi: 10.1080/02671520600942388

Cullen, J. B., Jacob, B. A., & Levitt, S. D. (2005). The impact of school choice on student outcomes: an analysis of the Chicago Public Schools. *Journal of Public Economics*, *89*(5-6), 729–760. doi: 10.1016/j.jpubeco.2004.05.001

Deming, D. J., Hastings, J. S., Kane, T. J., & Staiger, D. O. (2014). School choice, school quality, and postsecondary attainment. *American Economic Review*, *104*(3), 991-1013. doi: 10.1257/aer.104.3.991

Denton, N. A. (1995). The persistence of segregation: Links between residential segregation and school segregation. *Minnesota Law Review*, *80*, 795–824.

Eisenkopf, G. (2010). Peer effects, motivation, and learning. *Economics of Education Review, 29*(3), 364–374. doi: 10.1016/j.econedurev.2009.08.005

Epple, D., & Romano, R. E. (2011). Peer effects in education: A survey of the theory and evidence. In J. Benhabib, A. Bisin & M. O. Jackson (Eds.), *Handbook of social Economics Vol. 1* (pp. 1053–1163). Elsevier: North-Holland.

Fernández-Enguita, M. (2008). Escuela pública y privada en España: La segregación rampante. *RASE: Revista de la Asociación de Sociología de la Educación, 1*(2), 42–69.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education finance and policy*, *2*(4), 376-394. doi: 10.1162/edfp.2007.2.4.376

Frankel, D. M., & Volij, O. (2011). Measuring school segregation. *Journal of Economic Theory, 146*(1), 1–38. doi: 10.1016/j.jet.2010.10.008

Gibbons, S., Silva, O., & Weinhardt, F. (2017). Neighbourhood turnover and teenage attainment. *Journal of the European Economic Association*, *15*(4), 746–783. doi: 10.1093/jeea/jvw018

Gibbons, S., & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in England. *Oxford Bulletin of Economics and Statistics*, *78*(4), 548–575. doi: 10.1111/obes.12095

Givord, P. (2019a). Does greater social diversity in schools have an impact on equity in learning outcomes? *PISA in focus, 97*, 1–6. doi: 10.1787/35d2ab24-en

Givord, P. (2019b). How are school-choice policies related to social diversity in schools? *PISA in Focus*, *96*, 1–6. doi: 10.1787/2d448c77-en

Glazerman, S., & Dotter, D. (2017). Market signals: Evidence on the determinants and consequences of school choice from a citywide lottery. *Educational Evaluation and Policy Analysis*, *39*(4), 593-619. doi: 10.3102/0162373717702964

González-Betancor, S. M., & López-Puig, A. J. (2016). Grade Retention in Primary Education Is Associated with Quarter of Birth and Socioeconomic Status. *PLOS ONE, 11*(11), e0166431. doi: 10.1371/journal.pone.0166431

González-Betancor, S. M., & Marcenaro-Gutiérrez, O. D. (2018). Equidad y rendimiento académico: un delicado binomio a nivel regional. *Presupuesto y Gasto Público*, *90*, 117–134. ISSN 0210-5977.

Gorard, S., and Taylor, C. (2002). What is segregation? A comparison of measures in terms of 'strong'and 'weak'compositional invariance. *Sociology*, *36*(4), 875-895. doi: 10.1177/003803850203600405

Gortázar, L., Mayor, D. & Montalbán, J. (2019). *School Choice, Students Mobility and School Segregation: Evidence from Madrid.* I Southern Spain Workshop on Economics of Education, Málaga.

Goux, D., & Maurin, E. (2007). Close neighbours matter: Neighbourhood effects on early performance at school. *The Economic Journal, 117*(523), 1193–1215. doi: 10.1111/j.1468-0297.2007.02079.x

Guinea-Martín, D., Mora, R., & Ruiz-Castillo, J. (2015). The joint effect of ethnicity and gender on occupational segregation. An approach based on the Mutual Information Index. *Social science research*, *49*, 167–178. doi: 10.1016/j.ssresearch.2014.08.007

Gutiérrez, G., Jerrim, J., & Torres, R. (2019). School Segregation Across the World: Has Any Progress Been Made in Reducing the Separation of the Rich from the Poor? *The Journal of Economic Inequality*, *18,* 157–179. doi: 10.1007/s10888-019-09437-3

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of applied econometrics*, *18*(5), 527–544. doi: 10.1002/jae.741

Harding, D. J. (2003). Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy. *American Journal of Sociology, 109*, 676–719. doi: 10.1086/379217

Harris, D., & Williams, J. (2012). The association of classroom interactions, year group and social class. *British Educational Research Journal*, *38*(3), 373–397. doi: 10.1080/01411926.2010.548547

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. doi: 123/4/1373/1933172

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review*, *90*(5), 1209–1238. doi: 10.1257/aer.90.5.1209

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. doi: 10.1016/j.jpubeco.2005.11.002

Jacob, W. J., & Holsinger, D. B. (2008). Inequality in education: A critical analysis. In D. B. Holsinger, & W. J. Jacob (Eds.), *Inequality in education* (pp. 1–33). Dordrecht: Springer. doi: 10.1007/978-90-481-2652-1_1

Jerrim, J., Lopez-Agudo, L. A., & Marcenaro-Gutierrez, O. D. (2020). Posh but poor. The association between relative socio-economic status and children's academic performance. *Review of Income and Wealth*, in press. doi: 10.1111/ roiw.12476

Jeynes, W. (2000). Assessing school choice: A balanced perspective. *Cambridge Journal of Education*, *30*(2), 223–241. doi: 10.1080/03057640050075206

Klugman, J., & Lee, J. C. (2019). Social closure, school socioeconomic composition, and inequality in college enrollments. *Social science research*, *80*, 156–185. doi: 10.1016/j.ssresearch.2018.12.021

Krüger, N. (2018). An Evaluation of the Intensity and Impacts of Socio-economic School Segregation in Argentina. In Bonal, X. & Bellei, C. (Eds.), *Understanding School Segregation: Patterns, Causes and Consequences of Spatial Inequalities in Education* (pp. 103–122). London: Bloomsbury Academic.

Lavy, V., Paserman, M. D., & Schlosser, A. (2011). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, *122*(559), 208–237. doi: 10.1111/j.1468-0297.2011.02463.x

Lavy, V., Silva, O., & Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics*, *30*(2), 367–414. doi: 10.1086/663592

Lindbom, A. (2010). School choice in Sweden: Effects on student performance, school costs, and segregation. *Scandinavian journal of educational research*, *54*(6), 615–630. doi: 10.1080/00313831.2010.522849

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science*, *337*(6101), 1505–1510. doi: 10.1126/science.1224648

Mancebón-Torrubia, M. J., & Pérez-Ximénez de Embún, D. (2009). Segregación escolar en el sistema educativo español. Un análisis a partir de PISA 2006. *Investigaciones de Economía de la Educación,* 2009, vol. 4, p. 63–77. ISBN: 978-84-96221-37-6.

Mancebón-Torrubia, M. J., & Pérez-Ximénez de Embún, D. (2010). Una valoración del grado de segregación socioeconómica existente en el sistema educativo español. Un análisis por comunidades autónomas a partir de PISA 2006. *Regional and Sectoral Economic Studies*, *10*(3), 129–148.

Martínez, L., & Ferrer, A. (2018). *Mézclate conmigo: De la segregación socioeconómica a la educación inclusiva.* Madrid: Save the Children, España.

Martins, L., & Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review, 29*(6), 1016–1033. doi: 10.1016/j.econedurev.2010.05.001

Massey, D. S. (2012). Reflections on the dimensions of segregation. *Social Forces*, *91*(1), 39–43. doi: 10.1093/sf/sos118

MECD (2011). *Evaluación general de diagnóstico 2010. Educación secundaria obligatoria. Segundo curso. Informe de resultados*. Madrid: Ministerio de Educación.

Mora, R., & Ruiz-Castillo, J. (2011). Entropy-based segregation indices. *Sociological Methodology*, *41*(1), 159–194. doi: 10.1111/j.1467-9531.2011.01237.x

Murillo, F. J., Belavi, G., & Rodríguez, L. M. P. (2018). Segregación escolar público-privada en España. *Papers. Revista de Sociologia*, *1*(1), 1–31. doi: 10.5565/rev/papers.2392

Murillo, F. J., & Martínez-Garrido, C. M. (2018). Magnitud de la Segregación escolar por nivel socioeconómico en España y sus Comunidades Autónomas y comparación con los países de la Unión Europea. *Revista de la Asociación de Sociología de la Educación (RASE)*, *11*(1), 37–58. doi: 10.7203/RASE.11.1.10129

Musset, P. (2012). *School Choice and Equity: Current Policies in OECD Countries and a Literature Review*. OECD Education Working Papers, No. 66. Paris: OECD Publishing. doi: 10.1787/5k9fq23507vc-en

Musterd, S., Marcińczak, S., Van Ham, M., & Tammaru, T. (2017). Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. *Urban Geography*, *38*(7), 1062–1083. doi: 10.1080/02723638.2016.1228371

Musterd, S., Van Gent, W. P., Das, M., & Latten, J. (2016). Adaptive behaviour in urban space: Residential mobility in response to social distance. *Urban Studies*, *53*(2), 227–246. doi: 10.1177/0042098014562344

Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal, 50*(4), 714–754. doi: 10.3102/0002831213481240

Palardy, G. J. (2015). High school socioeconomic composition and college choice: multilevel mediation via organizational habitus, school practices, peer and staff

attitudes. *School Effectiveness and School Improvement*, *26*(3), 329–353. doi: 10.1080/09243453.2014.965182

Pont, B., Figueroa, D. T., Albiser, E., Wittenberg, D., Zapata, J., & Fraccola, S. (2014). *Education Policy Outlook: Spain*. Paris: OECD Publishing.

Prieto-Latorre, C., Marcenaro-Gutiérrez, O., & Vignoles, A. (2020). School segregation in public and semiprivate schools primary schools in Andalusia. British Journal of Educational Studies, in press. doi: 10.1080/00071005.2020.1795078

Schleicher, A., Zimmer, K., Evans, J., & Clements, N. (2009). *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD Publishing.

Shakeel, M., Anderson, K., & Wolf, P. (2016). The participant effects of private school vouchers across the globe: A meta-analytic and systematic review. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2777633

Simon, F., Małgorzata, K., & Pont, B. (2007). *Education and training policy no more failures ten steps to equity in education: Ten steps to equity in education*. Paris: OECD Publishing.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, *75*(3), 417–453. doi: 10.3102/00346543075003417

Söderström, M., & Uusitalo, R. (2010). School choice and segregation: evidence from an admission reform. *Scandinavian Journal of Economics*, *112*(1), 55–76. doi: 10.1111/j.1467-9442.2009.01594.x

Stacy, C., Meixell, B., & Srini, T. (2019). Inequality Versus Inclusion in US Cities. *Social Indicators Research*, *145*(1), 117-156. doi: 10.1007/s11205-019-02090-3

Stasz, C., & Von Stolk, C. (2007). *The use of lottery systems in school admissions*. RAND Working Paper WR-460-SUT, 1–18.

Statistics of the Regional Government of Andalusia (2012). *Alumnado escolarizado en el sistema educativo andaluz*. Junta de Andalucía: Conserjería de Educación y Deporte. https://www.juntadeandalucia.es/organismos/educacionydeporte/servicios/estadistica-cartografia.html

Theil, H., & Finizza A. J. (1992). A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts. In B. Raj, & J. Koerts (Eds.), *Henri Theil's Contributions to Economics and Econometrics. Advanced Studies in Theoretical and Applied Econometrics, vol. 25* (pp. 1030–1040). Dordrecht: Springer.

UNESCO (2018). *Handbook on measuring equity in education*. Montreal: UNESCO Institute for Statistics.

Vardardottir, A. (2013). Peer effects and academic achievement: a regression discontinuity approach. *Economics of Education Review, 36*, 108–121. doi: 10.1016/j.econedurev.2013.06.011

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review*, *76*(5), 713–736. doi: 10.1177/0003122411420816

Appendix

Table A1. Criteria which regulate the school admission of Andalusian students

Criterion	Points allocated/assigned
Place of residence or parental working place is located in the catchment area of the high school	10 points
Place of residence or working place is located in the boundary catchment area of the high school	6 points
Siblings attending high school	6 points
Family structure	
-Large family or single parent family	2 points
Parents work in the high school	1 point
Annual per capita family income	
-Under 25% of the IPREM (under 1,863.79 euros)	2 points
-From 25% to 33.33% of the IPREM (from 1,863.79 euros to 2,485.05 euros)	1.5 points
-From 33.33% to 50% of the IPREM (from 2,485.05 euros to 3,727.57 euros)	1 point
-From 50% to 66.67% of the IPREM (from 3,727 euros to 4,970.09 euros)	0.5 points
Existence of a disability	
-Student' disability	2 points
-Parents' disability	1 point
-Siblings' disability	0.5 points

Note: IPREM (*Indicador Público de Renta de Efectos Múltiples*) is the reference indicator for determining social benefits in Spain. In 2011 and 2012 this indicator is 7,455.14 euros.
Source: Authors' own elaboration drawn up on the basis of legislation (Law 53/2007, of 20th February 2007).

Table A2. Descriptive statistics by number of schools within the catchment area (at school level)

		Schools located in catchment areas with 1 school		Schools located in catchment areas with more than 1 school	
		Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Index of socioeconomic and cultural level		-0.382	0.355	-0.099	0.623
Number of students		50.833	25.735	47.461	31.586
Proportion of students from the quartile of ESCS distribution	**1st quartile**	0.364	0.136	0.282	0.201
	2nd quartile	0.302	0.091	0.260	0.119
	3rd quartile	0.210	0.085	0.239	0.111
	4th quartile	0.124	0.096	0.219	0.194
School funding	**Proportion of public schools**	0.995	0.071	0.695	0.460
	Proportion of semiprivate schools	0.005	0.071	0.305	0.460
Proportion of fathers with the level of education in the school	**Incomplete primary education or did not attend school**	0.243	0.119	0.177	0.157
	EGB or Compulsory Secondary Education	0.437	0.108	0.374	0.166
	Medium grade formation course	0.171	0.080	0.202	0.106
	High grade formation course	0.077	0.054	0.099	0.074
	University degree, PhD	0.072	0.074	0.149	0.163
Proportion of mothers with the level of education in the school	**Incomplete primary education or did not attend school**	0.183	0.104	0.145	0.134
	EGB or Compulsory Secondary Education	0.502	0.125	0.408	0.179
	Medium grade formation course	0.164	0.079	0.203	0.105
	High grade formation course	0.063	0.052	0.090	0.067
	University degree, PhD	0.088	0.080	0.154	0.153
Schools which are located in a municipality with	**Fewer than 5,000 inhabitants**	0.522	0.501	0.157	0.364
	From 5,000 to 10,000 inhabitants	0.370	0.484	0.064	0.245
	From 10,000 to 50,000 inhabitants	0.076	0.266	0.299	0.458
	From 50,000 to 100,000 inhabitants	0.022	0.146	0.114	0.318
	More than 100,000 inhabitants	0.011	0.104	0.365	0.482
Number of schools (observations)		198		1,297	

Note: The number of observations of size of the municipality of catchment areas with 1 school is 184 (instead of 198) because of missing values in this variable. "Std. Dev." stands for "standard deviation".
Source: Authors' own elaboration.

Table A3. Descriptive statistics by catchment area

		Mean	Std. Dev.
Index of socioeconomic and cultural level		-0.160	0.410
Number of schools within the catchment area		4.649	3.322
Number of schools within the catchment area	**Public**	3.233	1.679
	Semiprivate	1.416	2.346
Proportion of schools within the catchment area	**Public**	0.798	0.246
	Semiprivate	0.202	0.246
Proportion of parents with the level of education in the catchment area	**Incomplete primary education or did not attend school**	0.105	0.061
	EGB or Compulsory Secondary Education	0.374	0.118
	Medium grade formation course	0.224	0.053
	High grade formation course	0.114	0.044
	University degree, PhD	0.183	0.118
Catchment areas which are located in a municipality with	**Fewer than 5,000 inhabitants**	0.240	0.428
	From 5,000 to 10,000 inhabitants	0.118	0.324
	From 10,000 to 50,000 inhabitants	0.358	0.480
	From 50,000 to 100,000 inhabitants	0.075	0.264
	More than 100,000 inhabitants	0.208	0.407
Number of catchment areas (observations)		279	

Notes: "Std. Dev." stands for "standard deviation".
Source: Authors' own elaboration.

Table A4. Descriptive statistics by type of school (at school level)

		Public schools		Semiprivate schools	
		Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Index of socioeconomic and cultural level		-0.300	0.498	0.361	0.635
Number of students		49.039	33.995	43.858	24.904
Proportion of students from the quartile of ESCS distribution	**1st quartile**	0.339	0.189	0.153	0.165
	2nd quartile	0.282	0.113	0.207	0.116
	3rd quartile	0.224	0.109	0.272	0.107
	4th quartile	0.154	0.133	0.368	0.226
Proportion of fathers with the level of education in the school	**Incomplete primary education or did not attend school**	0.212	0.158	0.097	0.124
	EGB or Compulsory Secondary Education	0.415	0.156	0.278	0.149
	Medium grade formation course	0.186	0.107	0.238	0.094
	High grade formation course	0.087	0.070	0.126	0.077
	University degree, PhD	0.100	0.109	0.260	0.204
Proportion of mothers with the level of education in the school	**Incomplete primary education or did not attend school**	0.172	0.134	0.084	0.111
	EGB or Compulsory Secondary Education	0.460	0.162	0.289	0.158
	Medium grade formation course	0.183	0.101	0.249	0.100
	High grade formation course	0.077	0.064	0.118	0.067
	University degree, PhD	0.108	0.105	0.259	0.190
Schools which are located in a municipality with	**Fewer than 5,000 inhabitants**	0.223	0.416	0.008	0.087
	From 5,000 to 10,000 inhabitants	0.079	0.269	0.030	0.172
	From 10,000 to 50,000 inhabitants	0.334	0.472	0.220	0.415
	From 50,000 to 100,000 inhabitants	0.110	0.313	0.124	0.330
	More than 100,000 inhabitants	0.255	0.436	0.618	0.487
Number of schools (observations)		902		395	

Note: Catchment areas with only 1 school are excluded. “Std. Dev.” stands for “standard deviation”.
Source: Authors’ own elaboration.

Table A5. Academic achievement of poor and rich students (first quartile and fourth quartile of the ESCS distribution, respectively) according to the socioeconomic and cultural composition of the school they attend (schools from the first and fourth quartile of the ESCS distribution)

		Schools from the 1st quartile (25% of the *poorest* schools)			Schools from the 4th quartile (25% of the *richest* schools)		
	Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Students from 1st quartile (25% of the *poorest* students)	**Standardised scores in reading**	4,343	-0.495	1.008	1,156	-0.133	0.924
	Standardised scores in maths	4,446	-0.447	0.905	1,140	-0.180	0.891
Students from the 4th quartile (25% of the *richest* students)	**Standardised scores in reading**	593	0.331	0.949	8,671	0.625	0.742
	Standardised scores in maths	602	0.287	1.013	8,765	0.632	0.902

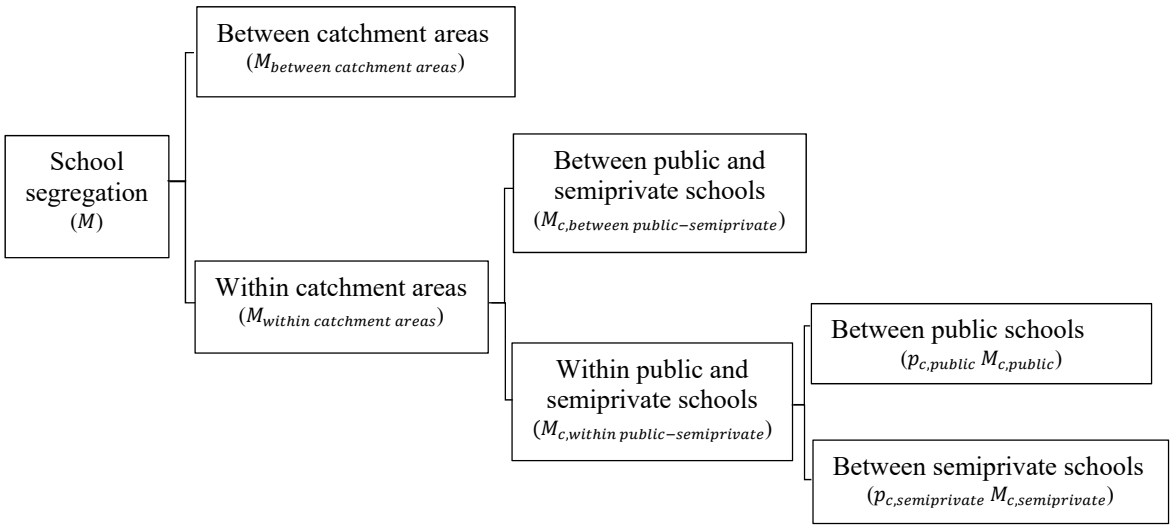
Note: Students who attend schools located in catchment areas with only 1 school are excluded. Mean comparison tests have been performed and the difference in academic achievement between schools from the 1st quartile of the ESCS distribution and the 4th quartile is statistically significant. "Obs." stands for "Observations" and "Std. Dev." stands for "standard deviation".
Source: Authors' own elaboration.

Table A6. Academic achievement of poor and rich students (first quartile and fourth quartile of the ESCS distribution, respectively) according to the socioeconomic and cultural level of the catchment area (catchment areas from the first and fourth quartile of the ESCS distribution)

		Catchment areas from the 1st quartile (25% of the *poorest* catchment areas)			**Catchment areas from the 4th quartile (25% of the *richest* catchment areas)**		
	Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Students from 1st quartile (25% of the *poorest* students)	**Standardised scores in reading**	2,857	-0.418	1.015	3,517	-0.289	0.962
	Standardised scores in maths	2,948	-0.417	0.899	3,522	-0.313	0.903
Students from the 4th quartile (25% of the *richest* students)	**Standardised scores in reading**	683	0.412	0.907	9,850	0.584	0.777
	Standardised scores in maths	686	0.412	0.987	9,952	0.598	0.924

Note: Students who attend schools located in catchment areas with only 1 school are excluded. Mean comparison tests have been performed and the difference in academic achievement between catchment areas from the 1st quartile of the ESCS distribution and the 4th quartile is statistically significant. "Obs." stands for "Observations" and "Std. Dev." stands for "standard deviation".
Source: Authors' own elaboration.

Figure A1. Decomposition of school segregation



Source: Authors' own elaboration.

Online Supplemental Material

Table S1. Segregation of students by quartiles of the ESCS distribution. Subsample of catchment areas with different types of schools (132 catchment areas which have both public and semiprivate schools). Absolute values

	Segregation among 1st, 2nd, 3rd, 4th quartiles	**1st quartile vs {2nd, 3rd, 4th quartiles}**	**2nd quartile vs {1st, 3rd, 4th quartiles}**	**3rd quartile vs {1st, 2nd, 4th quartiles}**	**4th quartile vs {1st, 2nd, 3rd quartiles}**
Total: School segregation	0.185	0.081	0.036	0.022	0.115
Between catchment areas	0.076	0.031	0.014	0.006	0.053
Within catchment areas	0.109	0.050	0.022	0.016	0.062
• **Between public and semiprivate schools**	0.034	0.017	0.005	0.003	0.020
• **Within public and semiprivate schools**	0.075	0.033	0.017	0.013	0.042
o **Between public schools**	0.039	0.019	0.007	0.007	0.020
o **Between semiprivate schools**	0.036	0.014	0.010	0.006	0.022

Notes: The segregation of public and semiprivate schools within a catchment area is weighted by the proportion of students attending public and semiprivate schools, respectively. The percentage of students attending public schools is 61.65% and the percentage of students attending semiprivate schools is 38.35%.
Source: Authors' own elaboration.

Table S2. Segregation of students by quartiles of the ESCS distribution. Subsample of catchment areas with different types of schools (132 catchment areas which have both public and semiprivate schools). Percentage decomposition by column

	Segregation among 1st, 2nd, 3rd, 4th quartiles	**1st quartile vs {2nd, 3rd, 4th quartiles}**	**2nd quartile vs {1st, 3rd, 4th quartiles}**	**3rd quartile vs {1st, 2nd, 4th quartiles}**	**4th quartile vs {1st, 2nd, 3rd quartiles}**
Total: School segregation	100.00%	100.00%	100.00%	100.00%	100.00%
Between catchment areas	41.08%	45.45%	42.42%	27.27%	46.09%
Within catchment areas	58.92%	54.55%	57.58%	72.73%	53.91%
• **Between public and semiprivate schools**	18.38% [A]31.19%	15.58% [A]34.00%	12.12% [A]22.73%	13.64% [A]18.75%	17.39% [A]32.26%
• **Within public and semiprivate schools**	40.54% [A]68.81%	38.96% [A]66.00%	45.45% [A]77.27%	59.09% [A]81.25%	36.52% [A]67.74%
o **Between public schools**	21.08% [B]52.00%	25.97% [B]57.58%	24.24% [B]41.18%	31.82% [B]53.85%	17.39% [B]47.62%
o **Between semiprivate schools**	19.46% [B]48.00%	12.99% [B]42.42%	21.21% [B]58.82%	27.27% [B]46.15%	19.13% [B]52.38%

Note: In general, the reference level to compute percentages is the total segregation between schools (calculated using the first row of Table S1, "Total: School segregation"). There are some exceptions: [A] The reference level is segregation within catchment areas for each column (third row of Table S1). [B] The reference level is segregation within the same type of school within a catchment area for each column (fifth row of Table S1).
Source: Authors' own elaboration.

Table S3. Segregation of students by quartiles of the ESCS distribution. Subsample of catchment areas with 1 school. Absolute values

	Segregation among 1st, 2nd, 3rd, 4th quartiles	**1st quartile vs {2nd, 3rd, 4th quartiles}**	**2nd quartile vs {1st, 3rd, 4th quartiles}**	**3rd quartile vs {1st, 2nd, 4th quartiles}**	**4th quartile vs {1st, 2nd, 3rd quartiles}**
Total: School segregation	0.091	0.041	0.017	0.021	0.042

Source: Authors' own elaboration.