UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E ATUÁRIA
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Racial gaps in intergenerational mobility and school factors: evidence from the State of São Paulo

Desigualdades raciais na mobilidade intergeracional e fatores escolares: evidências do Estado de São Paulo

Bruna Dominici Cricci
Orientador: Prof. Dr. Ricardo de Abreu Madeira

São Paulo
2022

BRUNA DOMINICI CRICCI

Racial gaps in intergenerational mobility and school factors: evidence from the State of São Paulo

Desigualdades raciais na mobilidade intergeracional e fatores escolares: evidências do Estado de São Paulo

Dissertação apresentada ao Programa de Pós-Graduação em Economia do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Ricardo de Abreu Madeira

Versão Original

São Paulo

2022

Acknowledgements

Gostaria de expressar meu especial agradecimento e apreço ao meu orientador, professor Dr. Ricardo de Abreu Madeira, por ter sido um mentor excelente não só nessa árdua trajetória que é a produção de um artigo científico, como também por seus conselhos sobre minha carreira profissional. Agradeço pela sua paciência, transmissão de conhecimento, por incentivar a produção da minha pesquisa e me permitir crescer como pesquisadora. Agradeço ao professor Dr. Marcos de Almeida Rangel, pelas excelentes contribuições e comentários para esta pesquisa.

Aos meus pais, por me darem força e coragem para seguir meus sonhos, por me proporcionarem os meios necessários para concluir essa etapa e por todo apoio durante esse processo. Ao meu querido irmão, Marcelo, por toda companhia, parceria e irmandade. Aos meus avós, por todo apoio e carinho.

Aos meus queridos amigos da USP, por tornarem essa trajetória mais leve e feliz. À minha grande parceira e amiga, Isadora, por ser um exemplo de mulher forte e inteligente em um ambiente majoritariamente masculino, como também por toda sua empatia, amor e acolhimento durante esse processo. Aos meus amigos, Caio, Matheus e José, por todas as risadas, desabafos e trocas acadêmicas e profissionais.

Um agradecimento especial às minhas amigas, mulheres que andam de mãos dadas comigo durante toda a minha jornada, por acreditarem no meu potencial, quando eu mesma não acreditei, por me darem colo, amor e muita força e coragem para seguir essa trajetória.

Por fim, gostaria de agradecer aos membros do meu comitê de dissertação pelo tempo e interesse que dedicaram a este projeto. E à Faculdade de Economia e Administração da Universidade de São Paulo e aos seus colaboradores pela sua excelência educacional.

O presente trabalho foi realizado com apoio da Fundação ao Instituto de Pesquisas Econômicas (Fipe) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradeço a ambas as organizações por seu apoio financeiro.

Abstract

This research paper explores unique data from the state of São Paulo, Brazil, that allows documenting how educational factors relate to intergenerational mobility for different racial groups. Conjunctly, we discuss how these factors are associated with the racial mobility gap. First, we find a strong intergenerational persistence of income and a significant difference in mobility between whites and blacks, possibly leading to persistent disparities across generations. Secondly, we document that after the 9th grade, the income trajectory of low and high-income students begins to diverge considerably. In addition, having a high school diploma and attending schools with good performance on the high-stakes test are associated with higher upward mobility rates for all students, but the wealthier seem to appropriate more of the benefits of school performance. These two educational factors are also related to smaller racial gaps for upward mobility and remaining in wealth, but they do not seem able to prevent blacks from falling socially or from the poverty trap. Furthermore, the data reveal that students from all social strata have better outcomes when they are among wealthier peers. However, schools with a high concentration of high-income parents have a more significant racial gap. We also find that students who attend more racially diverse schools have higher expected future earnings as young adults, regardless of race. However, racial integration in schools exclusively seems insufficient to close the existing racial gap. Finally, we discover that students with higher test scores than their peers are associated with higher future absolute gains and smaller racial gaps.

Key words: Intergenerational Mobility, Inequality, Races, Education, Brazil

JEL Classification: J62, D63, J15, I21

Resumo

Este artigo explora dados únicos do estado de São Paulo, que permitem documentar como fatores educacionais e escolares se associam à mobilidade intergeracional para diferentes grupos raciais e como estes fatores se relacionam com as disparidades raciais na mobilidade. Primeiro, encontramos uma forte persistência intergeracional de renda e uma diferença significativa entre indivíduos brancos e pretos, possivelmente levando a maiores disparidades entre gerações. Em segundo lugar, documentamos que após o 9º ano, a trajetória de renda de estudantes de baixa e alta renda começa a divergir consideravelmente. Além disso, verificamos que ter um diploma do ensino médio e estudar em escolas com alto desempenho na prova do SARESP estão associados a maiores taxas de mobilidade ascendente para todos os alunos. Contudo, aqueles de famílias mais abastadas parecem se apropriar mais dos benefícios advindos de escolas com maiores notas. Estes dois fatores educacionais também estão relacionados a menores diferenças raciais para mobilidade ascendente e permanência na riqueza, mas não parecem prevenir indivíduos pretos da queda de classe social ou da "armadilha da pobreza". Os dados revelam também que estudantes de todos os estratos sociais têm melhores salários futuros quando estão entre pares de famílias de classes mais altas. Entretanto, essa categoria escolar possui uma maior diferença racial. Também descobrimos que estudantes que freqüentam escolas mais diversas racialmente têm maiores rendimentos futuros esperados quando jovens adultos. Entretanto, essa diversidade racial sozinha parece ser insuficiente para fechar a lacuna racial existente na mobilidade. Finalmente, descobrimos que os estudantes com maior notas na escola em comparação com seus pares estão associados a maiores ganhos absolutos futuros e a menores diferenças raciais.

Palavras-chave: Mobilidade intergeracional, Desigualdade, Raças, Educação, Brasil

Códigos JEL: J62, D63, J15, I21

List of Figures

List of Tables

Contents

1 Introduction

During the last 20 years, Brazil has ranked as the seventh largest economy in the world, has lifted about 25 million people out of poverty[1], and has considerably improved access to education. However, it has never experienced an effective reduction in social inequality. According to the World Inequality Report (from Chancel et al. (2022)), available estimates suggest that the top 10% of income share has always been higher than 50% of total national income. Currently, the country ranks as the ninth most unequal country in the world, with a Gini index of 0.53. This persistent inequality has multiple causes, which may or may not be an intergenerational inheritance, that is, the transmission of social disadvantages between generations within the same family. Moreover, it is widely known that one of the characteristics of social inequality in Brazil is racial inequality (Telles (2014)). Therefore, studying intergenerational mobility, its associated factors, and how these affect mobility's racial disparities is relevant from an equity perspective.

In recent decades, many researchers have focused on trying to estimate the degree to which parents' income determines children's opportunities and its associated factors. However, the lack of adequate data has limited the scope of this type of research. The challenge is to have data that makes it possible to link income across more than one family generation, for a large number of individuals, together with their observable characteristics, to allow assessing associated factors. Recently the estimation of the degree of intergenerational income mobility has been overcome in the US by Chetty et al. (2014b) and in Brazil by Britto et al. (2022) by combining administrative data at the individual level. However, the factor's association still needs to be improved since these databases do not have a very high level of detail concerning individuals' observable characteristics. Understanding these associated factors is crucial for designing effective policies to fight these inequalities.

In this research paper, therefore, we explore a unique data by creating a new database for the State of São Paulo that allows documenting how school and

[1]World Bank data from the National Sample Survey of Households (PNAD/IBGE) and the data presented refers to the years 2003 through 2014.

educational factors relate to intergenerational mobility for different racial groups. That is, our analysis is complementary to Chetty et al. (2014b) and Britto et al. (2022). At the individual level, we investigate how educational and school characteristics (such as school performance, educational trajectory, social and racial segregation in schools, and student relative performance to their peers) are associated with the intergenerational mobility of blacks and whites. Additionally, we discuss how these factors seem to be associated with the racial mobility gap.

To the best of our knowledge, this is the first research to document in detail how educational and school factors relate to intergenerational mobility when comparing different racial groups in a highly unequal context such as Brazil. Although this study does not propose causal relationships, it brings interesting discussions by relating literature on education and race to assess intergenerational mobility.

The structuring of the databases for this research is inspired by recent methods used in literature that combine administrative data. We merge two educational databases from the State of São Paulo, the System for Evaluating School Performance in the State of São Paulo (SARESP) with the enrollment register of schools in the same state, with a labor market's data also of São Paulo, the Annual Report of Social Information (RAIS). This match allows us to connect longitudinally two generations, parents and children, besides allowing the observation of the child's educational variables. From SARESP, we estimate the parents' permanent income, using the Principal Component Analysis (PCA) method. SARESP is a high-stake exam that, besides examining the students, surveys their parents, asking about their socioeconomic conditions, with questions related to goods, services and assets that the family owns. From this same database, we obtained the race variable from the students' self-reported race and the associated educational and school factors. From RAIS, we take information on the children's salaries already in the labor market. The enrollment register of schools is used to make it possible to connect the information from these two other databases and to track student's educational trajectories. Finally, we merge the databases at the individual level based on the student's full name and date of birth.

We divided our empirical analysis into two estimation methods from the recent intergenerational mobility literature for each associated school factor. First, we estimate the rank-rank regression for each school subgroup. Following Chetty et al. (2014b), we analyze the coefficients of absolute mobility, which measures the mean rank of a child born in the lowest income percentile, and the relative mobility, which associates the mean percentile rank of children and their parents' income ranks. For example, we compare these coefficients between students studying in high-scoring schools and those in low-scoring schools. Ultimately, these coefficients will show the different degrees of income persistence among students from different educational backgrounds.

In the second part, we estimate a transition matrix between parental and child income quintiles for each school group, but now also stratifying by race. These matrices allow us to analyze the movement of white and black children, separately, of ancestry, descent, or social permanence by quintile of the income distribution, with the social stratum occupied by their parents as the starting point reference. This analysis shows us how the probabilities of upward and downward mobility, poverty trap, or remaining in wealth are distributed among different racial groups and educational backgrounds.

Next, using the matrices' probabilities, we create a racial gap indicator: the ratio between the white's and the black's probability. This indicator informs us which educational context favors more each racial group. For example, we can see whether racially diverse schools benefit more blacks to ascend socially than racially segregated schools. In this research, we focus on the black's perspective, highlighting which school categories analyzed seem to favor blacks more. We do this because it is known that racial disparities are persistent and pronounced in Brazil. Among many other areas, these differences are present in the education and schooling of individuals (Madeira & Rangel (2014)), as well as in the labor market and remuneration (Firpo et al. (2021)). The recent racial literature brings evidence that some of this difference stems from discrimination (Botelho et al. (2015); Bertrand & Mullainathan (2004); Quillian et al. (2017)). Our research, therefore, dialogues with this racial inequality literature, trying to understand whether or not specific aspects of inequality are associated with social characteristics. As well as analyzing which forces in the educational system could narrow the gap of intergenerational mobility between blacks and whites in the long-run.

Previous to the school results, we begin presenting the estimates for our three working samples and dividing them by race. Evidence shows a strong intergenerational income persistence in São Paulo: an individual born into a more socially advantaged family is, on average, two times more likely to be among the top 25% of earners as a young adult. This persistence also occurs in maintaining individuals in poverty. An individual born into a low-income family is, on average, 1.5 times less likely to remain in poverty than individuals born in the highest quintile of the income distribution. In addition, we find significant differences between blacks and whites. Blacks are 1.6 percentage points (p.p.) more likely to experience downward mobility, 2.1 p.p. less likely to experience upward mobility, 0.51 p.p. more likely to remain in poverty, and 5.6 p.p. less likely to remain rich. Consequently, under the assumption that mobility rates remain constant across generations, these findings show that the disparity between blacks and whites will evolve across generations, which is aligned with Chetty et al. (2014b) and Britto et al. (2022) findings.

These estimations aren't necessarily new insights for the literature, as it has already been done by Chetty et al. (2014b) and Chetty et al. (2020) in the US and Britto et al. (2022) in Brazil. Therefore, we only present it to put our results in perspective with those of recent literature. The numbers we find are smaller than those of Britto et al. (2022) for Brazil's population and Chetty et al. (2014b) and Chetty et al. (2020) for the US's population. These differences occur due to significant differences between the data used. In this regard, we discuss the main differences between our research and theirs, which has to do with the nature of our data. Although they work with representative populations from Brazil and the US, we have to work with a subpopulation. We do this in order to get more granular and detailed school information at the individual level. Thus, the comparison between the parameters of this study and theirs should be made with caution.

As mentioned, on the other hand, this research paper brings innovation to the literature on intergenerational mobility by linking it to educational and school factors. Our first finding in this regard shows that 9th-graders have higher relative intergenerational mobility than 12th-graders, suggesting that what occurs between these grades seems to have a significant impact on relative immobility. High school is a determining moment in the student's future income, and the choices made during this period seem to be very much associated with the parent's socioeconomic background. Thus, this seems to be a critical time to guarantee compensatory policies that give low-income individuals future income trajectories more similar to the wealthy.

Our evidence also shows that students who attend schools with good performance in a high-stakes exam have higher absolute intergenerational mobility. However, this higher mobility is not uniformly distributed since the difference increases across the parental income rank. Students from high-income families seem to appropriate the benefits of school performance more. There are a few explanations for this phenomenon, one would be that students coming from low-income families face higher opportunity costs of studying due, for example, to credit constraints (Lochner & Monge-Naranjo (2012); Flug et al. (1998)) or because school quality is complementary to family human capital (Brunello & Checchi (2005)). This suggests that other mechanisms trap the individual in poverty or cause them to persist in poverty for generations.

This finding relates to a poverty trap literature, dialoguing with the idea that a high-performance school alone does not seem to be a sufficient mechanism to lift the individual out of poverty or at least to end relative differences. Several mechanisms can maintain individuals in the poverty trap, such as dysfunctional institutions, neighborhood effects (Durlauf et al. (2006)) or, for example, economies and coordination failures, irregular nutrition and investments, and behavioral and geographic poverty traps (Kraay & McKenzie (2014)).

Regarding racial differences, we also find that, in relative terms, high-test scores schools favor blacks for upward mobility and remaining rich compared to low-test scores schools. This finding corroborates with Arias et al. (2004), who show that equalizing access to quality education is one of the critical factors for reducing racial income inequality in Brazil. Despite that, an intriguing result of ours is that schools with better test scores fail to prevent blacks from falling socially or keeping them in poverty.

Students who concluded high school have higher absolute and relative mobilities than those who drop out. This increase in absolute mobility is associated primarily with higher upward mobility rates. In addition, finishing high school seems to favor blacks relative to whites in keeping them in wealth and for social ascendance than for dropout students, but this does not occur in other socioeconomic settings. In other words, having a high school diploma reduces the racial gap in intergenerational mobility for upward mobility or among the wealthiest. However, it does not prevent blacks from downward mobility or poverty traps. This finding goes in line with Chetty et al. (2020), who show that high school dropout rates cannot directly explain the sharp disparities observed in outcomes, between blacks and whites, at younger ages.

Students from all social strata have better outcomes when they are among wealthier peers. We have found that schools with a higher concentration of wealthy parents are associated with higher absolute mobility than schools with a higher concentration of low-income parents (16% lower) or the same proportion of high and low-income parents (10% lower). This finding is consistent with Chetty et al. (2022), who showed that communities in which people with low SES interact less with people with high SES exhibit less upward income mobility across generations. Moreover, the degree of friendship between people with low and high SES is strongly associated with upward income mobility (Chetty et al. (2022)). The economic connections literature shows that economic mobility can be facilitated by connections with people who can provide access to information and employment opportunities or with people who can help shape aspirations (Loury (1977) and Lin & Dumin (1986)). On the other hand, schools with a higher concentration of high-income families are associated with more significant racial mobility gaps than the other categories (socially diverse schools and schools with more families from lower social classes).

Pupils from racially diverse schools have higher expected future earnings as young adults, regardless of race. However, racial integration exclusively seems insufficient to close the existing racial gap because although they favor blacks in cases of upward mobility and getting out of poverty, they do not favor them in cases of downward mobility and staying in wealth. In this case, school segregation can be considered a good proxy for neighborhood segregation since the dominant rule in São Paulo regarding which school the children will attend is based on the house's proximity. Therefore, these results can be associated with the residential segregation literature. In this sense, our results dialogue with those of Chetty et al. (2020), who show that only reducing residential segregation is insufficient to close the black-white gap.

The last educational factor studied is the student's getting good test scores in their school environment. We find that students who get good grades compared to their peers have higher absolute mobilities across the entire parental income distribution. In this analysis, we bring a recent discussion that examines the rank effect in school, showing that students' relative position affects subsequent academic performance, and some mechanisms that mediate this rank effect are related to students' motivations and aspirations (Ladant & Bargagli-Stoffi (2022); Elsner & Isphording (2018)). We can connect our results to this literature discussion by speculating that students' mechanisms of self-confidence and aspirations may also be affecting intergenerational mobility. We also find that being a relatively good student can reduce racial disparities.

Ultimately, this paper aims to contribute to academia by bringing new stylized facts to the literature of intergenerational mobility and racial inequality that the causal inference literature may yet identify the causes behind. Also, we seek to bring new insights to the educational public policy debate by specifying which factors may be associated with greater intergenerational mobility in a highly unequal context such as Brazil. In the field of race studies, this work aims to make more explicit the intergenerational inequalities between blacks and whites in a country where structural racism is deeply rooted in its economic and social structures. Conjunctly, we seek to contribute to the racial debate by elucidating which school factors are associated with less racial disparities. The results could assist policymakers that aim to tackle racial inequalities for future generations through educational-focused public policies.

2 Estimating Wealth

This section presents the methodology and the step-by-step for constructing the parents' Wealth Index. We obtain the parent's information from the SARESP's database. SARESP is a high-stake exam applied to students from public schools in São Paulo (more details about this database in Section 4). In this exam, the student's parents must answer a socioeconomic survey that requests information regarding the family's services and assets in their homes. Therefore, we summarize this information into a wealth indicator using the Principal Component Analysis methodology. The advantage of using a socioeconomic indicator is that it reflects the families' long-term wealth and is not subject to short-term fluctuations.

We have organized this section into two parts. The first part presents the Principal Component Analysis methodology, and the second describes the step-by-step for the indicator's estimation.

Principal Component Analysis Method.

We create a wealth index using the Principal Component Analysis methodology to estimate the parents' wealth. This methodology consists in a "multivariate statistical technique used to reduce the number of variables in a data set into a smaller number of dimensions" (Vyas & Kumaranayake (2006)). From a set of correlated variables, PCA extracts a set of uncorrelated "principal components". Each principal component is a weighted linear combination of the original variables. Mathematically, suppose we have n correlated variables X1 to Xn. In that case, each principal component is the sum of each variable multiplied by its weight (the weight for each variable is different in each principal component), i.e.:

$$
\begin{aligned}
PC_1 &= a_{11}X_1 + a_{12}X_2 + ... + a_{1n}X_n \\
&\vdots \\
PC_m &= a_{m1}X_1 + a_{m2}X_2 + ... + a_{mn}X_n
\end{aligned}
$$

Where a_{mn} represents the weight for the nth principal component and the nth variable.

The weights are given by the correlation matrix's eigenvectors or the covariance matrix's (in case the original data has been standardized). The eigenvalue of the corresponding eigenvector gives the variance for each principal component. The number of variables in the PCA determines how many principal components we will have. The components are ordered so that the first principal component (PC_1) explains the data's most significant variation, subject to the constraint that the sum of the squared weights is equal to one. Therefore, we assume that this first principal component represents wealth index, or the socioeconomic position indicator. In other words, the first component is a latent measure of socioeconomic status, the linear combination that accounts for the largest proportion of variance that is common to all items (Torche (2015)).

The other components (PC_2, ... PC_n) are not correlated with each other, nor with the first, by construction. Therefore, each following component explains additional variation in the data but less than the first component, subject to the same constraint. It is worth noting that the higher the degree of correlation between the original variables in the data, the fewer components are needed to capture standard information.

Wealth Index Estimation.

With this methodology, we created the parent's wealth index using the SARESP socioeconomic survey the student's parents answered. For its estimation, we follow the step-by-step approach presented by Vyas & Kumaranayake (2006) and Torche (2015). First is the selection of asset variables.

In order to select items to be included in the index, we considered the entire set of goods and services asked in the survey: daily newspaper and general information magazine, dictionary, piped water, sewage, electric light, piped gas, garbage collection service, color television (working or being repaired), radio (working or being repaired - excluding car radio), car, vacuum cleaner (working or being repaired), washing machine (working or being repaired), VCR and/or DVD (working or being repaired), refrigerator (working or being repaired), freezer (stand-alone appliance or part of a duplex refrigerator), telephone, computer, cable, satellite or pay TV, microwave, internet. We also considered whether the individual hires a maid service (monthly and working at least from Monday to Friday) and the house condition payment, that is, whether it is owned (paying or already paid) or not.

We transform these items into binary variables, and each frequency is analyzed. The items with more than 95% frequency or less than 5% will not be considered in the indicator since they are weak indicators of socioeconomic position (see Table 1). In other words, variables with low standard deviations would have low weight in the PCA and are, therefore, of little use in differentiating socioeconomic levels. Thus, in this first step, a descriptive analysis is conducted for all items in the questionnaire, analyzing the means and standard deviation.

The items selected using the frequency criterion were: daily newspaper and general information magazine, dictionary, sewage, piped gas, garbage collection service, radio, car, vacuum cleaner, washing machine, VCR and/or DVD, refrigerator, freezer, telephone, computer, cable, satellite or pay TV, microwave, internet, maid service, and house condition payment. Therefore, the variables removed were TV, piped water, and electric light, which are variables that less than 5% of the study population does not possess.

A significant concern when using this method is the "ability of the index to discriminate across the entire socioeconomic structure in the context being analyzed, including the lower and upper ends" (Torche (2015)). This discrimination is achieved by including items that distinguish access to resources among the poor and rich. In our case, note that the collection services and refrigerators distinguish among the poors, and maid service and vacuum cleaners among the wealthy.

Next, we estimate the socioeconomic index using a few methods - the PCA with correlation matrix, the PCA with covariance matrix, and Factor Analysis. The three indicators' correlations are greater than 0.95 (see Appendix's Table A.1), so we understand that any of the three methods will lead to similar results. We follow using the indicator estimated by PCA and covariance matrix because it had a higher proportion of the first component, and according to Vyas & Kumaranayake (2006), "if the raw data has been standardized, then PCA should use the covariance matrix". And in our case, the variables have been transformed into binary.

After running the PCA, the first component is extracted. Next, the value of the first component was predicted for each individual in our three samples. Finally, the predicted value is used as a measure of socioeconomic condition and for creating the parents' percentile rank (the next Section (3) will explain in more detail about the rankings).

This socioeconomic indicator (also called "Asset Index" by the literature) is a measure that reflects the families' long-term wealth. Torche (2015) shows how different authors present different names for this same concept. "For example, Filmer & Pritchett (2001) indicate that the index captures household wealth, which they then interpret as a proxy for long-run economic status or expenditures, McKenzie (2005) refers to living standards, Ferguson et al. (2003) mention permanent income, and Sahn & Stifel (2003) speak of well-being". This indicator reflects the household's wealth and is therefore not subject to temporary or short-term fluctuations, such as income.

Table 1 – Parents good and services considered in the construction of the Wealth Index

9th grade Students			12th grade Students	
	Mean	Sd	Mean	Sd
Maid Service	7,0%	0.26	7,0%	0.26
Piped Gas	21,0%	0.41	23,0%	0.42
Vacuum Cleaner	30,0%	0.46	32,0%	0.47
Daily Newspaper & Magazine	39,0%	0.49	38,0%	0.49
Freezer	41,0%	0.49	49,0%	0.50
Cable, Satellite or pay TV	43,0%	0.49	46,0%	0.50
Internet	53,0%	0.50	59,0%	0.49
Car	54,0%	0.50	56,0%	0.50
Telephone	59,0%	0.49	60,0%	0.49
Microwave	61,0%	0.49	64,0%	0.48
Computer	64,0%	0.48	69,0%	0.46
Owned House (paid or paying)	65,0%	0.48	65,0%	0.48
Washing Machine	83,0%	0.38	84,0%	0.37
Radio	86,0%	0.34	87,0%	0.34
Sewage	87,0%	0.33	89,0%	0.32
DVD or VCR	88,0%	0.33	89,0%	0.32
Dictionary	90,0%	0.30	91,0%	0.29
Garbage Colletion Service	92,0%	0.26	93,0%	0.26
Refrigerator	95,0%	0.22	92,0%	0.28
Piped Water	97,0%	0.18	97,0%	0.17
Color Television	97,0%	0.18	97,0%	0.17
Eletric Light	98,0%	0.14	98,0%	0.13

Notes: Descriptive statistics (mean and standard deviation) of the entire set of goods and services asked in the SARESP's survey by grade. The mean also corresponds to the frequency of individuals who own the respective good (or service).

3 Intergenerational mobility calculation

This section presents the conceptual framework of the two methods we use to calculate intergenerational mobility. First, we present the rank-rank regression method, which makes an association between children's and fathers' income rankings. And then, the transition matrices method, which documents the children's movement along the income quintiles, given their parents' position in the income distribution.

Rank-rank regression.

This method is used in the intergenerational mobility literature (Chetty et al. (2020); Chetty et al. (2014b); Dahl & DeLeire (2008); Britto et al. (2022))) and consists of a simple relationship between the children's income ranking and the father's income ranking. The relationship between both variables tends to be linear, therefore, it is possible to model it such that the estimated parameters can be used comparatively across groups. More specifically, we estimate the following regression:

$$y_{i,t} = \alpha_g + \beta_g y_{i,t-1} + \epsilon_{i,t} \tag{1}$$

where i is the family index, and t is the generation, so $\mathrm{y}_{i,t}$ denotes the children's income percentile of family i of generation t, and $\mathrm{y}_{i,t-1}$ is the parent's income percentile of family i and generation *t-1*. Furthermore, $\epsilon_{i,t}$ is an idiosyncratic shock independent across generations and has expectation $\mathrm{E}[\epsilon_{i,t}] = 0$. Let also *g(i)* denote the group under analysis, which in this study will be both races and students coming from different school backgrounds.

We can interpret $\alpha_g \in [0,1]$ as the absolute rank mobility, that is the mean rank for a child of group g, whose parents have income rank $\mathrm{y}_{i,t-1} = 0$. The parameter $\beta_g \in [0,\ 1]$ measures the rate of relative mobility, that is, the relation between the mean percentile rank of children and their parents' income ranks for group g. It is worth noting that for absolute mobility, the higher the α, the higher the absolute mobility of that population. On the other hand, in the case of relative mobility, the higher the β, the lower the relative mobility since a higher β implies a greater distance between individuals who are one percentile apart in their parents' income distribution. In a society of perfect intergenerational mobility, the rank-rank slope should be zero so that the parent's income ranking would not affect the children's income ranking.

Therefore, α and β parameters point to the expected ranking of a child born into a family from any point in the income distribution. Moreover, such parameters can be compared across groups. For instance, it is possible to reach the expected outcomes of black children compared to white children or the expected outcome of children who attended high-performance schools compared to the ones in worse-performance schools, etc. These comparisons will constitute a significant part of this paper. Through this, we can provide guidance on types of intervention and policies that may be more effective in reducing these disparities.

Transition Matrix.

This second approach is also used in the intergenerational mobility literature (Corak & Heisz (1999); Chetty et al. (2014b)). This method divides the population of children and parents into n equal-sized quantiles and ranked in income order. Then, it presents the parent and child distributions across these quantiles, analyzing the movement between them over a generation. This approach can be considered complementary to rank-rank regression providing more detailed information than regression. More specifically, an advantage of the transition matrix is that it shows the direction of mobility, that is, the behavior of social ascendent and descendent per quantile of the social distribution.

Take the joint distribution (Y_i, X_i), where Y_i refers to the children's income and X_i to the parent's income. In large samples, one can characterize the joint distribution of (Y_i, X_i) nonparametrically, and we provide this in the form of a 5x5 transition matrix, that is, by ranking the parents and children by quintiles. In this study, we will mainly analyze the four extremes of the matrix, i.e.: (i) the probability that a child born to parents in the bottom quintile remains in this quintile; (ii) the probability that a child born to parents in the bottom quintile rises to the top quintile of the distribution; (iii) the probability that a child born to parents in the top quintile falls to the bottom quintile; and (iv) the probability that a child born to parents in the top quintile remains in this same quintile.

Furthermore, we calculate the racial gap for each quintile to analyze the differences between races. This gap is simply characterized by the white and black probability ratio per quintile. Therefore, the higher this ratio, the more significant the difference between whites' and blacks' upward or downward mobility.

4 Data

For the primary analyses, we combined three databases: (i) the School Performance Evaluation System of the State of São Paulo (SARESP) (2008, 2010 to 2012); (ii) the Enrollment Database of São Paulo's State (2009 to 2014); and (iii) the Formal Labor Market Data of Brazil (RAIS) (2007 to 2018). The union of these databases allows us to monitor several socioeconomic indicators of individuals from school years to the labor market. In addition, and for some complementary analyses, we also use the National Sample Survey of Households (PNAD/IBGE) (2016 to 2019), but this data is not part of our main results.

Formal Labor Market Data Brazil - RAIS

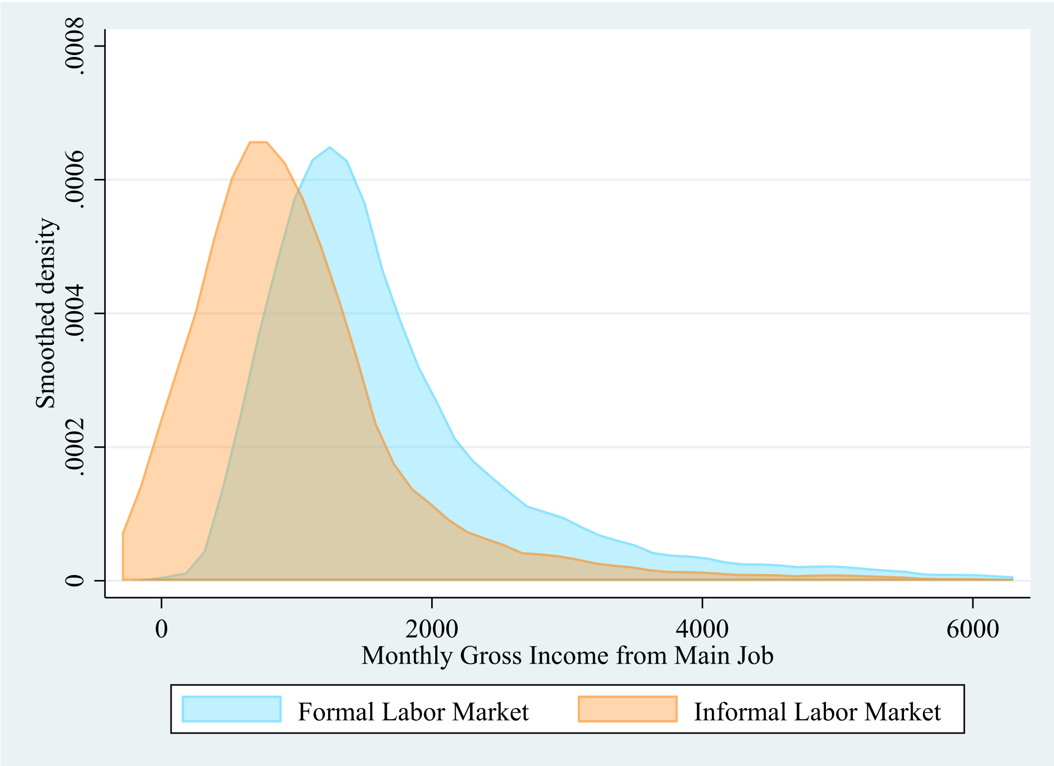
The RAIS database is an annual employer-employee report from the Brazilian Ministry of Labor and Employment. The annual report is made by employers and is mandatory according to Brazilian legislation. This base is a fundamentally important instrument for monitoring and characterizing the formal labor market throughout the country. However, this research will be restricted to data from the state of São Paulo, considering that the SARESP evaluation is applied only in this federative entity. The main variables will be gender, salary, full name, date of birth, age, and CPF (Individual Taxpayer Registration Number) of the individuals from 2007 to 2018.

The Brazilian formal labor market consists of 60% of the population. This number is 68.9% in São Paulo.[2] Figure 1 plots the wage distribution by formal and informal markets in Brazil. The figure shows that the common support between the formal and informal markets is significant. However, as expected, there are more individuals with lower earnings in the informal market than in the formal. That is, the data from the formal labor market (RAIS) do not include a portion of individuals in the lower part of the wage distribution. This underrepresentation of individuals from the lower tail generates a bias in our results, overestimating mobility. We will discuss this limitation further in section 5 (subsection "Comparisons

[2]According to Brazilian Institute of Geography and Statistics (IBGE) estimates, 2022.

with other articles in the literature").

Figure 1 – Salary distribution, by formal vs. informal market in Brazil



Notes: This figure plots the distribution of the monthly gross income from the individuals' main job by the formal and informal labor market. The data is from the National Continuous Household Sample Survey (PNAD Contínua), IBGE, from 2016 to 2019.

School Performance Evaluation System of the State of São Paulo - SARESP

The SARESP is an annual evaluation system (high-stakes exam) applied by the Secretary of Education of São Paulo State (SEE-SP)[3] to diagnose and monitor the evolution of state education. The test is based on the Item Response Theory (IRT) methodology and seeks to assess the student's proficiency at the end of each educational "cycle" - such as 3rd, 5th, 7th, 9th and 12th grades. For example, the test applied to 12th-grade students will assess how well the student has grasped the expected content at the end of high school.

[3] The Secretariat of Education of São Paulo's State (SEE-SP) is the largest agency in Brazil dedicated to educating young people and children. There are 5.4 thousand schools and approximately 3.5 million students.

These data are especially interesting because it has not only the students' individual scores in the test scores but also a set of questionnaires from the school managers, parents, and students. For this research, we will use variables such as students' ethnic self-declaration obtained from the student questionnaire and the families' socioeconomic conditions when the students took the standardized test scores from the parental questionnaire. Also, we will use the Portuguese and Mathematics scores from the test score and the school codes for the analyses made at the school level.

The SARESP test is administered every November, at the end of the school year in Brazil, to the 3rd, 5th, 7th, 9th and 12th grades in state government schools. In Brazil, 80% of the total students attend public schooling, and this number is 77.2% in São Paulo[4], an expressive fraction of the universe of students in Brazil and São Paulo. We used the data from 2008, 2010, to 2012 from the SARESP's questionnaires. Unfortunately, 2009 could not be utilized due to problems with the student identification variable (RA), which makes it impossible to match it with the other databases of the study. Although we have more years of the SARESP's questionnaires available, this is the range used once these are the years we have available for the enrollment database (see the following subsection for further information about the Enrollment Database).

Enrollment Database of São Paulo's State

The enrollment database of São Paulo's state contains all students enrolled in basic education in the state (private or public schools). However, we will use only public school students since they are the ones for whom we have information on socioeconomic conditions (from SARESP's questionnaires). We will use the variables at the student level, such as the school in which they were enrolled in the respective year, gender, date of birth, and their identifier (RA). The latter is the primary key that allows the match between the enrollment and SARESP's databases. We have this data from 2008 to 2012.

[4]According to data from the National Institute for Educational Research and Study and Executive Secretary (INEP), specifically, School Census 2019.

Matching Structure: RAIS + SARESP + Enrollment

One contribution of this paper is creating a new database that can share information on parents' socioeconomic conditions when the student was in school, their school characteristics, and their income information when they are older in the labor market. That is, a database that contains information on the income and wealth of two generations, parents and children.

Data Cleaning

First, we do the data cleaning for our three primary databases, RAIS, SARESP, and Enrollment. Figure 2 shows the cleaning step-by-step and the number of observations we lost in each step.

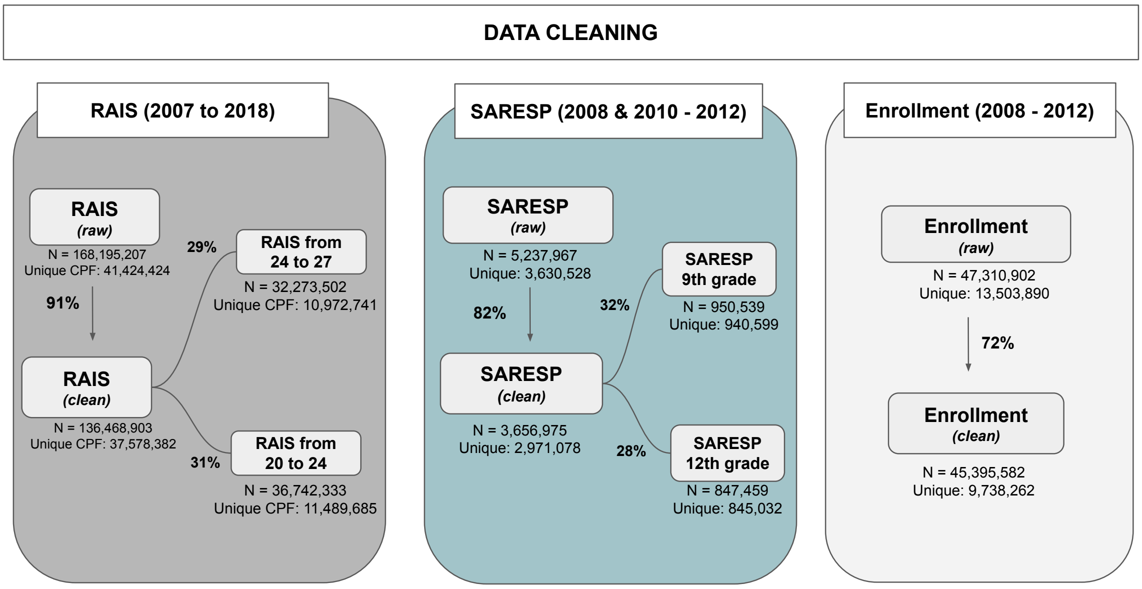
For RAIS, to mitigate the merge errors, we eliminate individuals with the same (full) name and birthdate but with a distinct identifier (CPF), and we restrict the sample to workers from 20 to 24 years old and from 24 to 27 based on their birth year. As shown in Figure 2, in the initial data cleaning, we get 91% of the individuals from the raw database, and each final RAIS sample represents around 30% of the cleaning database. It is worth noting that each final RAIS database has a unique person identifier (CPF), such that for the rows with repeated CPFs, we calculated the December median deflated earnings (see the subsection below "Variable definition" for more details).

In SARESP, we lost 18% of the individuals from the raw data because we only kept the following: (i) the students whose parents answered the socioeconomic questionnaire (that is, we discarded the ones in which grandparents or other responsible answered); and (ii) those who answered all the questions regarding the goods, services, and assets we will use to build the wealth index. Then, we restrict the sample to students in the 9th grade and the 12th grade. Both also represent about 30% of all students in the clean SARESP database, as shown in figure 2. Here, we also keep a unique identifier per person (RA), thus, individuals with more than one answer in the questionnaire are averaged over the answers before calculating the wealth index.[5]

[5] This is something that happened with low frequency. These are specific cases in which

Finally, in the enrollment database, also to mitigate errors, we eliminate individuals with the same (full) name and birthdate but with distinct RAs. In this process, we lost 28% of the individuals from the raw data.

Figure 2 – Diagram with the data cleaning step-by-step of each primary database



Merge 1: Link RA - CPF

The great challenge in joining the three databases is the inexistence of a common primary key. Therefore, first, we will have to create a database that connects the identifiers from both sides: RA, the student's identifier in SARESP, and CPF, the workers' identifier in RAIS. To do so, we will merge the enrollment database with RAIS.

In this first merge, we used the individual's full name and date of birth as the primary key. It is worth mentioning that Britto et al. (2022) merged RAIS with

students are enrolled in two schools because in one, they do regular high school, and in the other, they do technical education and take the SARESP exam more than once in the same year. Alternatively, there are cases in which we have students who did the 9th or 12th grade twice, so they answered the questionnaires two years in a row, for example.

the Cadastro Único (CadUnico) database using only individuals' full names. The paper shows that within states, 70% of the individuals have unique full names. This statistic was calculated taking into account only distinct full names. Therefore, in our case, the number is even higher since we consider the full name and birthdate.

Figure 3 shows we found 16% of the people from RAIS in SARESP. However, not all RAIS individuals are actually possible to find in SARESP, considering the years for which we have data available. Thus, in Figure 4, we filter only RAIS workers in the 1970-2000 birth cohorts since these are ages we can find both in school (from 2008 to 2012) and in the labor market (from 2007 to 2018). From this population of RAIS individuals, we found 74.3% of them in SARESP.

On the other hand, from the enrollment database, we found 62% of students in SARESP. It is important to remember that while the SARESP exam takes place only among students in São Paulo state public schools, the enrollment base contains all students enrolled in education (until high school) in the state. Thus, it is predicted that we would not find: students from the private education sector; students enrolled in 2013 and 2014 schools, since the SARESP data goes only until 2012; and students from municipal education.

Ultimately, this new database (we will call it intermediate database) will link the individual identifier RA with the other unique identifier, the CPF. Once this process is concluded, this new database contains all individuals from RAIS in São Paulo who also studied at school in São Paulo, with their respective identifiers: RA and CPF. People outside this database are not potential study subjects because they were not found in SARESP or RAIS. As shown in Figure 3, this base contains 21,051,506 observations and 5,989,823 unique CPF-RAs.

Figure 3 – Diagram of merge 1 that connects the identifiers (RA and CPF)

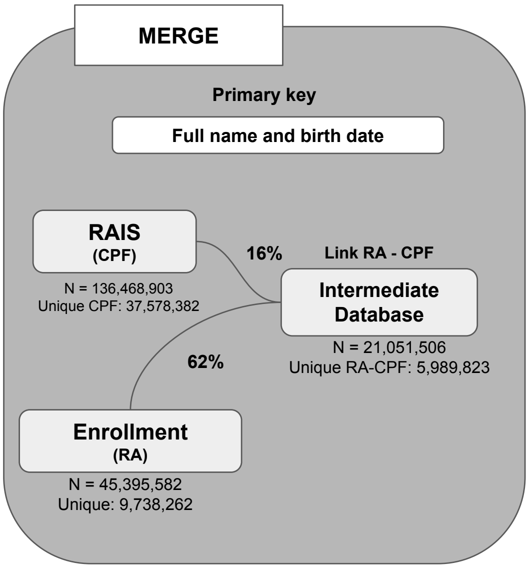
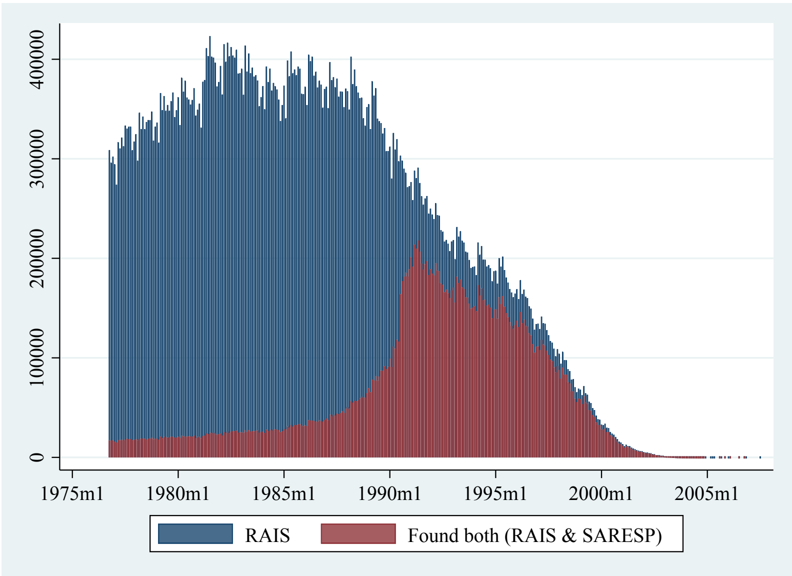


Figure 4 – Number of people in each database, by month and year of birth

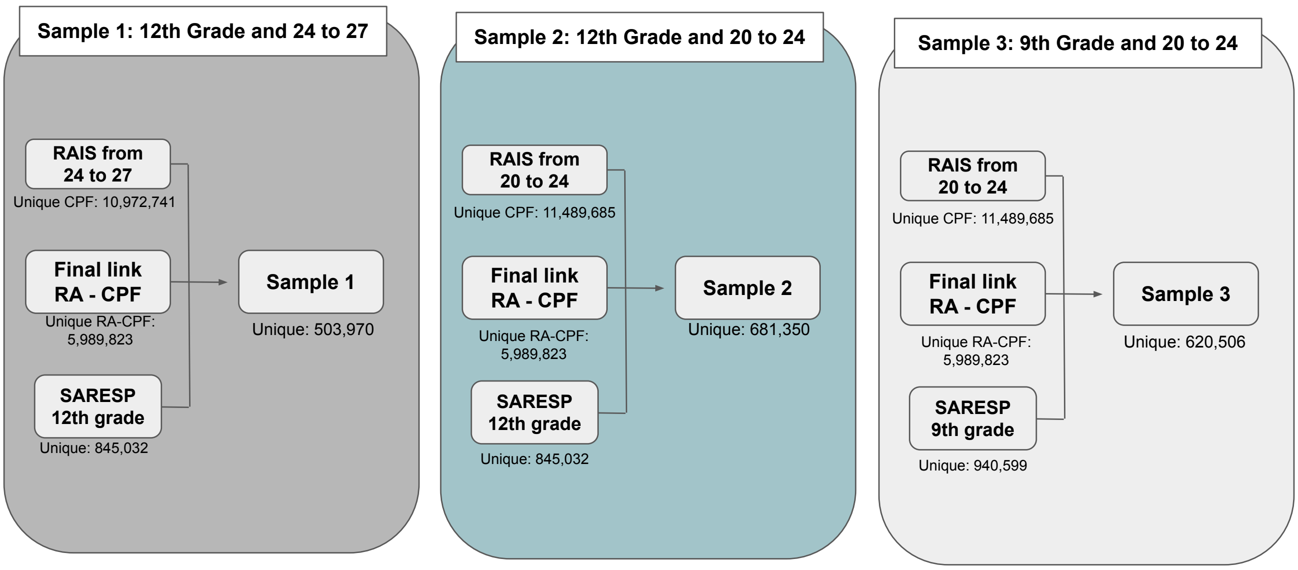


Notes: The X-axis represents the number of people in each database born from 1975 to 2005. Therefore, the blue bars represent the number of people in all years of RAIS that we have data available born in the respective month-years shown. Moreover, the red represents the number of people within this same birth interval found in our base that links RA-CPF.

Merges: Constructing the final samples

Following, we build our working samples. For all three, the procedure is the same, first combining the filtered RAIS with the intermediate base, which has the RA-CPF link, through the primary key CPF, then connecting this new base with the filtered SARESP through the RA key. As mentioned, the three databases have unique identifiers. Figure 5 shows the number of people found in both databases (RAIS and SARESP) and the number of individuals in our final three Working Samples.

Figure 5 – Diagram with the final merges to get to the working samples



Sample Definition

For our analysis, we will use three distinct samples. First, we take all 9th-grade students (from all the years that we have SARESP available) who were found in RAIS between the ages of 20 to 24. We chose this age range since they represent the ages with more 9th graders found in RAIS, which is consistent with the years we have both data available. In the SARESP database, we have the years 2008, 2010 to 2012, and in the RAIS database, 2007 to 2018. Therefore, by selecting students in the ninth grade or younger for the sample, we will only be able to find them still young in the labor market due to limited data availability.

In Brazil, there is a significant problem of students dropping out of school before reaching high school. Therefore, the 9th grade becomes an interesting and essential grade to analyze since it can bring us a perspective on how school performance factors can impact intergenerational mobility at a stage before high school, with a smaller selection bias than the one existing among 12th-grade students. Moreover, this grade makes it possible to analyze the relationship between students' school trajectories and intergenerational mobility.

Due to age restrictions in 9th-grade students, we work with a second sample composed of 12th-grade students (from all the years we have SARESP available) that were found in the RAIS between the ages of 24 and 27. This second sample allows us to analyze the students at a slightly older age in the labor market when the salary is closer to the long-term salary. The literature understands that it is around age 35 that the peak of individuals' labor return occurs (Murphy & Welch (1990)). So, although 24 to 27 years old is closer to 35, it is still not the recommended age in the literature. Therefore, we did an exercise to see the correlation between the salary ranking of individuals aged 20 to 27, the age we work with in our samples, with the salary ranking of individuals aged 33 to 40. Figure 9 shows a strong relationship between both rankings, of 0.63. Therefore, although we use a lower age than recommended by the literature, the individuals' wages at this age have high correlations with the wages at a more mature stage of the labor market.

Finally, and for comparison purposes with the 9th-grade exercise, our third sample represents the 12th-grade students of SARESP found between 20 and 24 years old in RAIS.

Tables 2 to 4, we show descriptive statistics comparing the RAIS and SARESP populations from the primary database with our working samples. Tables 2 and 3 present the variation of the student's test scores and the socioeconomic indicator along the distribution for 12th-grade and 9th-grade students, divided by race. The working samples have slightly better test scores and socioeconomic indicators than the initial population. The existing difference corroborates with the selection created by our merge, which selects only students working in the formal labor market, who possibly had better test scores and came from better socioeconomic conditions (see section 5, subsection "Comparisons with other articles in the literature" for more of this discussion).

Table 4 presents variations in the wage distribution in the working samples compared to the total population of workers. In this case, the table is not divided by race since we used this variable from the SARESP self-declaration. In short, this table also show that our samples are similar to the total population, with marginal differences.

Table 4 – Comparison between the working samples and RAIS population

	Sample 1	**Workers (24 to 27 years)**	**Sample 2**	**Sample 3**	**Workers (20 to 24 years)**
Mean	2,106.5	2,232.9	1,766.5	1,628.6	1,742.7
P25	1,414.4	1,335.2	1,309.6	1,278.7	1,244.5
P50	1,764.4	1,702.8	1,559.4	1,491.0	1,496.7
P75	2,420.6	2,473.7	1,974.4	1,805.8	1,911.9
P90	3,421.5	3,861.7	2,635.9	2,282.2	2,644.8
N	*443,216*	*29,697,332*	*591,660*	*532,838*	*33,803,596*

Notes: Sample 1 refers to parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS); Sample 2 to parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS); and Sample 3 refers to parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS). The wage is the December Deflated Earnings, using the consumer price index (IPCA, in Portuguese) with the base year of 2018.

Table 2 – Comparisons between the two working samples using 12th-grade students and the total population, by race

	Sample 1		**Sample 2**		**12th grade (all)**	
	White	Black	White	Black	White	Black
Panel A. Test Score						
Mean	278.2	267.8	277.6	267.4	275.6	264.8
P25	250.4	241.2	249.9	240.8	247.3	237.9
P50	276.8	265.6	276.3	265.2	273.8	262.3
P75	304.8	292.1	304.1	291.7	302.2	288.9
P90	329.9	317.2	328.8	316.6	327.9	314.4
N	*229,113*	*178,978*	*261,024*	*204,566*	*424,857*	*354,993*
Panel B. Wealth Index						
Mean	0.06	-0.19	0.06	-0.18	0.05	-0.19
P25	-0.58	-0.96	-0.57	-0.95	-0.59	-0.98
P50	0.28	-0.04	0.28	-0.03	0.28	-0.04
P75	0.74	0.56	0.74	0.57	0.75	0.56
P90	1.04	0.94	1.04	0.94	1.04	0.95
N	*245,302*	*192,084*	*261,024*	*204,566*	*344,078*	*268,003*

Notes: Sample 1 refers to parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS), and sample 2 to parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS). Test Score refers to the student's average test score in the Portuguese and Mathematics of SARESP. The primary database has all the students we had the information on grade and race in the raw SARESP. That is, before filtering only the students found in RAIS.

Figure 6 – Correlation between salary rankings at different ages

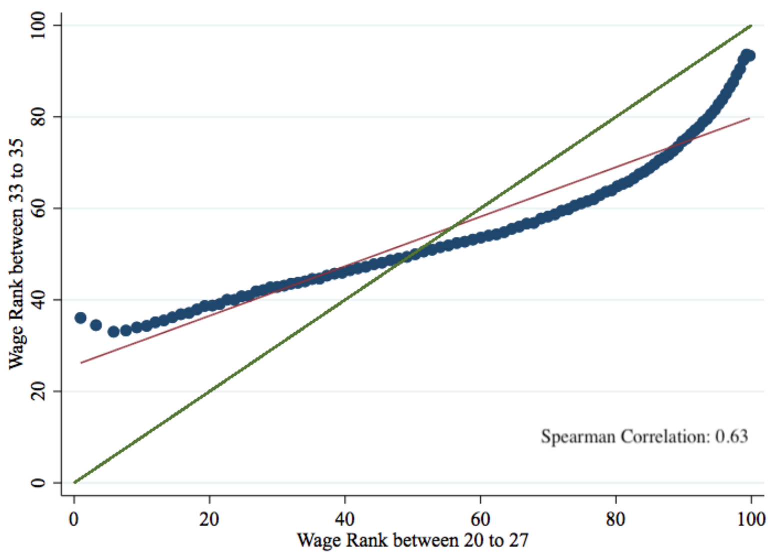


Table 3 – Comparison between the working samples using 9th-grade students and the total population, by race

	Sample 3		**12th grade (all)**	
	White	Black	White	Black
Panel A. Test Score				
Mean	248.6	237.8	245.4	234.0
P25	220.7	211.6	216.1	207.2
P50	247.4	235.4	243.4	230.9
P75	274.9	261.7	272.7	257.9
P90	299.5	286.1	298.8	283.5
N	*284,372*	*286,945*	*664,091*	*683,890*
Panel B. Wealth Index				
Mean	0.13	-0.13	0.14	-0.13
P25	-0.60	-0.89	-0.60	-0.90
P50	0.32	-0.09	0.34	-0.10
P75	0.86	0.65	0.89	0.65
P90	1.15	1.03	1.16	1.04
N	*293,030*	*296,486*	*443,693*	*438,918*

Notes: Sample 3 refers to parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS). Test Score refers to the student's average test score in the Portuguese and Mathematics of SARESP. The primary database has all the students we had the information on grade and race in the raw SARESP. That is, before filtering only the students found in RAIS.

Variable Definitions

This subsection will define and describe the variables used in our primary analyses.

Variable Definitions for Parents:

Wealth Index. Our measure for parents' wealth is the Wealth Index we created from the parents' SARESP socioeconomic questionnaire using Principal Component Analysis. In section 2, we explain in more detail the methodology used and the variables that compose this indicator. Table 5 presents the percentage of people who own each good, service and asset used in the index by quintile, that is, it shows one representative agent per quintile. For example, note that some items significantly differentiate an individual in the first quintile from an individual in the last quintile, such as the internet, computer, car, microwave oven, and freezer.

Table 5 – Percentage of people who own each good and service used in the Wealth Index, by quintile

Good & Services in the Wealth Index	**Quintile**				
	1	**2**	**3**	**4**	**5**
Maid Service	3,0%	5,0%	5,0%	7,0%	20,0%
Piped Gas	20,0%	24,0%	23,0%	29,0%	40,0%
Vacuum Cleaner	5,0%	19,0%	22,0%	41,0%	85,0%
Daily Newspaper & Magazine	28,0%	34,0%	34,0%	40,0%	56,0%
Freezer	32,0%	48,0%	54,0%	64,0%	80,0%
Cable, Satellite or pay TV	17,0%	31,0%	40,0%	55,0%	78,0%
Internet	0,9%	31,0%	86,0%	98,0%	99,0%
Car	24,0%	48,0%	53,0%	74,0%	93,0%
Telephone	27,0%	54,0%	67,0%	79,0%	93,0%
Microwave	24,0%	54,0%	67,0%	86,0%	98,0%
Computer	8,0%	63,0%	96,0%	99,0%	100,0%
Owned House (paid or paying)	55,0%	67,0%	66,0%	78,0%	93,0%
Washing Machine	61,0%	81,0%	89,0%	97,0%	99,0%
Radio	81,0%	87,0%	89,0%	94,0%	98,0%
Sewage	85,0%	90,0%	92,0%	93,0%	95,0%
DVD or VCR	75,0%	87,0%	91,0%	96,0%	99,0%
Dictionary	89,0%	93,0%	94,0%	95,0%	96,0%
Garbage Colletion Service	89,0%	92,0%	93,0%	94,0%	96,0%
Refrigerator	78,0%	84,0%	90,0%	95,0%	98,0%

Notes: To construct this table, we calculated the average by quintile of each binary variable that composes the Wealth Index.

Parent's Ranking. The parent's ranking is constructed based on the Wealth Index. For the rank-rank regression, we use the percentile rank, and for the transition matrix, we divided the samples into quintiles. We rank parents based on their Wealth Index relative to all other parents in 12th grade (for samples 1 and 2) or 9th grade (for Sample 3). That is, the ranking was not based only on the parents included in the final samples but on an external population - all parents of students in the analyzed school grade. Since our working samples represent a subpopulation of individuals, ranking only the closed population would not give us a good perspective on where the parents are classified in the total population of parents with children in the same school grade. Therefore, the ranking based on the external population gives a more realistic view of parent's socioeconomic status.

Variable Definitions for Children:

Income. Our main measure for children's income is the December median deflated earnings when children are between 20 and 24 (for the case of samples 2 and 3) and between 24 and 27 (for sample 1). The deflation was calculated using the consumer price index (IPCA, in Portuguese) with the base year of 2018.

Children's Ranking. The children's ranking is constructed based on the income specified above. For the rank-rank regression, we use the percentile rank, and for the transition matrix, we divide the populations into quintiles. As in the case of the parent's ranking, we create the ranking based on an external population, the RAIS's workers. That is, we rank children based on their income relative to all other workers aged 20 to 24 from RAIS (in the case of samples 2 and 3) and workers aged 24 to 27 from RAIS (in the case of sample 1). Since our working samples represent a subpopulation of individuals, ranking only the closed population would not give us a good perspective on where the children are classified in the total population of formal workers of their age. Therefore, the ranking based on the external population gives a more realistic view of children's socioeconomic status.

Race. We assign race to children using the information they report on the SARESP's students questionnaire, which is the student's self-declaration of race. We put in the "white" category those students who self-declared themselves as "White" in the questionnaire and the "black" category those who self-declared themselves as "Black" or "Pardo". "Pardo" is a race/colour category used by the Brazilian Institute of Geography and Statistics (IBGE) in Brazilian censuses that includes various shades of brown.

Racial Gap. For the transition matrix exercises, we created an indicator called Racial Gap. This indicator is calculated as the ratio between the white's probability of being in a respective cell of the transition matrix and the black's probability. Each cell represents one of the four socioeconomic situations - poverty trap, upward mobility, downward mobility, or remaining in wealth. The racial gap measures how much whites' probability of being in the respective socioeconomic situation is higher (> 1) or lower (< 1) than blacks'. However, note that the matrix has positive and negative economic situations, so this indicator should be interpreted cautiously. For instance, if we are talking about the poverty trap, an economically unfavorable scenario, and the indicator is less than 1, it means that whites have a lower probability than blacks of being in that position. Therefore, it is a scenario that favors whites more than blacks. On the other hand, if we analyze upward mobility, which is a desirable economic scenario, and the indicator is less than 1, it favors blacks more than whites.

However, our interpretation will not analyze each indicator separately; we will make a comparative analysis, verifying which school categories analyzed seem to favor blacks more. For example, for an upward mobility scenario, if we have two school categories, where Type 1 has a racial gap of 1.02, and Type 2 of 1.3, we will conclude that the Type 1 school is more favorable to blacks than Type 2 even if both have indicators greater than 1. In other words, this would be a scenario that is not narrowing the racial gap per se since whites' probability of upward mobility is still higher in both school categories. However, it indicates which of the two categories is better for blacks.

In this research paper, we will analyze the blacks' perspective. That is, we will always highlight when a situation is more favorable for blacks, as explained in the paragraph above. We do this because since blacks have higher chances of downward mobility and lower chances of upward mobility (more details in the "Blacks and Whites" subsection of the Results, Section 5), naturally, racial gaps in income will persist over time. Thus, by highlighting which educational settings the blacks are favored in, we ultimately indicate possible educational factors that favor narrowing racial gaps in the long run.

The indicator's interpretation we will follow throughout this study will be as follows:

- **For the Poverty Trap (P(Child in Q1| Parent in Q1)):** the **higher** the indicator, the more favorable the situation for blacks;
- **For Upward Mobility (P(Child in Q5| Parent in Q1)):** the **lower** the indicator, the more favorable the situation is for blacks;
- **For Downward Mobility (P(Child in Q1| Parent in Q5))** the **higher** the indicator, the more favorable the situation for blacks;
- **For Remaining Rich (P(Child in Q5| Parent in Q5)):** the **lower** the indicator, the more favorable the situation is for blacks.

Variable Definitions for School and Educational characteristics:

School Performance. To measure school performance, we use the Portuguese and Mathematics scores that 12th-grade (or 9th-grade) students took on the SARESP exam.

We intend to separate the effect of individual student performance on standardized tests from the school's performance as a whole. Therefore, for each student, we will calculate a different school average, which will be: the average score in Portuguese and Mathematics on the SARESP exam of all the students in this school minus the score of the respective student. Thus, the average will always be calculated with (n - 1) grades, where n is the number of school students in the 12th grade (or 9th grade, depending on the sample analyzed). Therefore, this measure will reflect the school performance without the student's grade, or in other words, the performance of the school according to the pool of students studying with that respective student, excluding his (or her) performance.

Educational Trajectory. To construct this variable, we used only Sample 3, which refers to 9th-grade students, since they are the ones that we can still have a trajectory until 12th grade. Therefore, we used the enrollment database, which gives us information about all years that the respective student was enrolled in a school. Note that 2014 is the last year we have the enrollment database available, thus, the only information we can use about 2014 is concerning students in 12th grade that year, otherwise, we have no way of knowing if that student continued in school or not in the following years. Note also that when we use "dropout students," it refers to those who have left school in the 9th grade or high school.

Below we describe how each of the two categories of this variable was constructed:

1. Students who dropped out - in this category, we only include students who were: (i) last found in the enrollment database in some year before 2014; and (ii) were enrolled in 9th, 10th or 11th grade that year.
2. High School Graduates - in this category, we included only students found in 12th grade in any year of the enrollment base.

Social Segregation in School. The construction of this categorical variable was based on another variable that measures social inequality in schools. The measure used was created at the school level, such that:

$$\textit{Social Inequality in Schools} = \frac{X}{Y},$$

where X = Number of students from the wealthiest quintile of the total population; and Y = Number of students from the lowest quintile of the total population.

The total population here is either 9th grade or 12th grade, depending on the sample. So, per school, we counted the number of students that belong to the highest quintile and the lowest quintile based on the parent's Wealth Index.

Next, we created the categorical variable, social concentration in schools, which has three categories:

1. Schools with a high concentration of high-income parents - represent the last quintile of the social inequality variable (top 20%);
2. Schools with a high concentration of low-income parents - represent the first quintile of the social inequality variable (bottom 20%);
3. Schools with a higher mix of low and high-income parents - represent the third quintile of the social inequality variable since it is in this quintile that

the variable crosses the 1 (same number of high and low-income families in the school);

Racial Segregation in School. The construction of this categorical variable follows the same logic of Social Segregation. Therefore, we created an intermediate variable at the school level such that:

$$\textit{Racial Inequality in Schools} = \frac{X}{Y},$$

where X = Number of white students in the respective grade (either 9th grade or 12th grade) and school; and Y = number of black students in the respective grade and school.

Next, we created the categorical variable, racial concentration in schools, which has three categories:

1. Schools with the highest concentration of whites - represents the highest decile of the social inequality variable (top 10%);
2. Schools with the highest concentration of blacks - represents the lowest decile of the social inequality variable (bottom 10%);
3. Schools with greater racial equality - represents the fourth decile of the social inequality variable since it is in this decile that the variable crosses the 1 (same proportion of whites and blacks in school).

Relatively good students. For this variable, we calculated the average between the Portuguese and Mathematics scores that each student took in the SARESP test in the respective grade (either 9th grade or 12th grade, depending on the sample), then we ranked the scores by school, and created a dummy that assumes one if the student belongs to the top 20% of students in that school in that grade, and zero if the student belongs to the bottom 20% of students in that school in that grade. This variable is a proxy for measuring whether a student is a relatively good-grade student or a relatively bad-grade student in their school context.

5 Results: Intergenerational Mobility

Full Sample

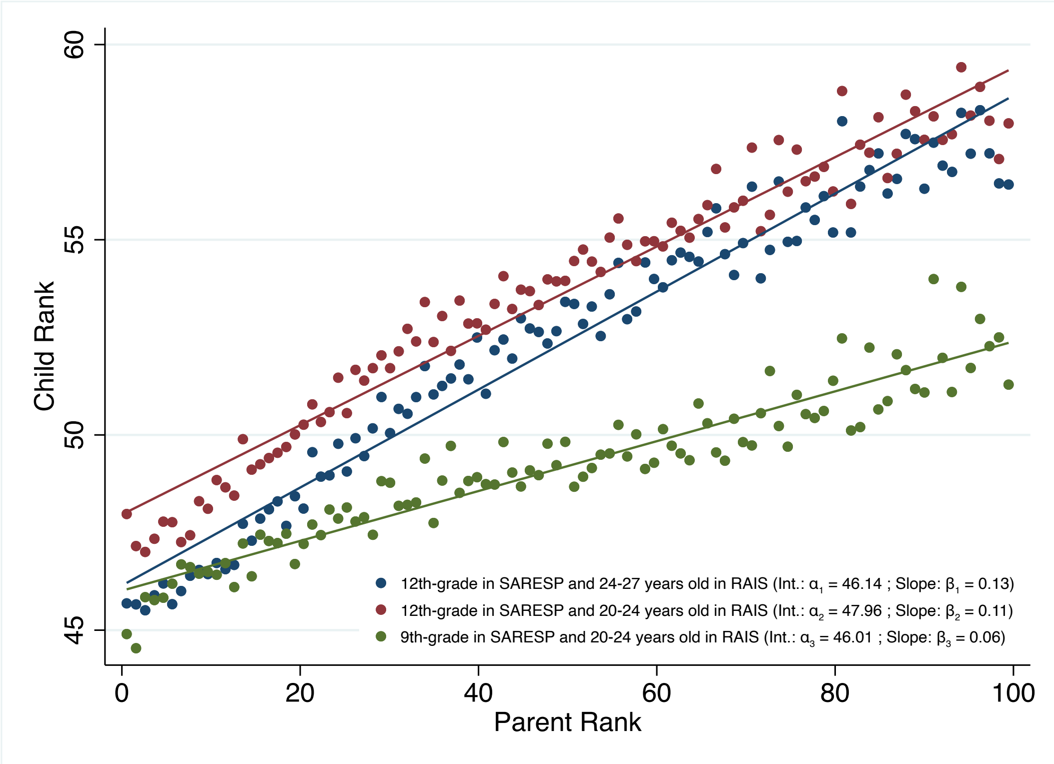
In this section, we present the intergenerational mobility estimations (α, β) using the framework introduced in Sections 2 and 3 for the three samples (see section 4, "Sample Definition", for more details regarding each sample):

- **Sample 1:** Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS);
- **Sample 2:** Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS);
- **Sample 3:** Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Following Chetty et al. (2014b), we rank the parents' socioeconomic status and the children's salary using percentile ranks. The parents' ranking was made based on their status relative to all other parents of 12th-grade or 9th-grade (respectively for each sample) students who took the SARESP test in the years we have data (2008, 2010-2012). Similarly, the children's salaries were ranked based on their earnings relative to all other children from 24 to 27 years old or from 20 to 24 (respectively for each sample) in the years we have RAIS's data (2007-2018).

Figure 7 shows intergenerational mobility for each sample separately, plotting the children's mean salary rank against their parents' mean income rank. As Chetty et al. (2014a) show, the rank-rank slope constitutes a simple way to summarize the degree of mobility in a society. Therefore, for each sample described above, the absolute and relative intergenerational mobility (α_i, β_i) was estimated based on specification 1 (see section 3).

Figure 7 – Intergenerational Mobility for each sample



Notes: Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS); Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS); Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

First, Figure 7 shows us a distinct pattern in the mobility of 9th graders compared to 12th graders, which justifies using three different samples in our analysis. Secondly, analyzing the relative mobility of Figure 7, we see a modest difference of 0.02 for the 12th-grade students because we measure children's incomes at slightly higher ages (20 to 24 vs. 24 to 27), reducing the life-cycle bias [6]. For Sample 1, we found a β_1 of 0.11 (in case of Sample 2 we have a $\beta_2 = 0.13$) means that a ten percentile increase in the parent's rank is associated with a 1.1 (or a 1.3) percentile increase in the children's rank, on average. On the other hand, the relative mobility for 9th-grade students (Sample 3) is higher ($\beta_3 = 0.06$) compared to

[6]The younger we measure the income of individuals, the more chances of a possible underestimation of intergenerational persistence in lifetime income because children with high lifetime income have steeper earnings profiles when young (Solon (1999); Grawe (2006)).

12th-grade students. This disparity probably occurs because, in 9th grade, we are analyzing the mobility of younger individuals, and there are still many possibilities concerning their future salaries. In other words, what happens in an individual's life from 9th grade to 12th grade greatly impacts relative immobility.

This figure suggests that while students are in 9th grade, the trajectory between individuals from low and high-income families is more similar (low relative mobility). However, after 9th grade, the parents' socioeconomic background becomes significantly relevant in determining their future income, and the trajectory of low and high-income students differs greatly. High school is a determining moment in the student's future income, and the choices made during this period seem to be very much associated with the parent's socioeconomic background. Thus, this seems to be a critical time to guarantee compensatory policies that give low-income individuals future income trajectories more similar to the wealthy.

It is possible to conjecture several factors in this period that differentiate the low and high-income students' trajectory. Considering the Brazilian context, we can speculate, for example, that students from wealthier families that attend public schools in 9th grade may have migrated to better-quality private schools during high school. Alternatively, even if the student remained in public school, the parents may have invested in a preparatory course for the Vestibular exam [7], which allowed them to enter better universities after high school. Further results will delve into the discussion of school differences between high and low-income students, as well as for different racial groups, and it will bring more insights to this discussion. Nevertheless, future research should be done to understand the critical aspects of low and high-income students' educational trajectories influencing this relative immobility during high school.

Table 6 report summary statistics for children and parents of each sample. The smallest sample we work with has 503 thousand students (from Sample 1). This sample is expected to be the smallest once it matches students from SARESP at the most advanced ages (24 to 27) in the labor market (RAIS). Furthermore, the sample for Sample 2 has 620 thousand students, and Sample 3 has 681 thousand.

[7] The Vestibular is a competitive exam high school graduate students undertake as a mandatory process to be accepted into Brazilian universities.

The median children's salary for Sample 1 is R$ 1,763.7, the highest median salary among the three samples. Since Sample 1 has only students that completed high school and were found when they were older in the labor market, it is also expected that this sample will have the highest salary. The median children's salary for Sample 2 is R$ 1,557.0, and for Sample 3 is R$ 1,490.5. Moreover, the mean parent income rank does not vary significantly between the samples, ranging from 48.7 in Sample 1 to 50.1 in Sample 2.

Secondly, Table 6 shows the intergenerational mobility of each sample, specifically parts of the transition matrices, filtering the quintiles 1 and 5 of each generation (the distribution's extreme). As mentioned in Section 3, transition matrices capture the degree of social mobility in society, as they present ascending/descending behavior in each quantile of the social distribution. This transition matrix reports how much the parents' socioeconomic condition is associated with the children's salary in the formal labor market for each sample when they are still young (20 to 24 years old or 24 to 27 years old, depending on the sample).

Analyzing Sample 2, for example, note that the probability that a son born in the bottom quintile moves to the top quintile is 14.8%. In comparison, the similar probability of a son born in the top quintile is 27.5%. This evidence a strong intergenerational income persistence in São Paulo as an individual born in a more socially favored family (upper quintile) is 2 times more likely to be in the top 25% of earnings as a young adult (20 to 24 years old). The interpretation of these probabilities for the other samples follows the same logic, therefore, for Sample 1, individuals born in the highest 25% of parent rank are 2.2 times more likely to remain there in comparison to individuals born in the bottom 25%, and this number is 1.7 for Sample 3. Intergenerational persistence also occurs in maintaining individuals in poverty. Again, looking at Sample 2 as an example, note that the probability of an individual born in a poverty trap (P(Q1|Q1)) remains there is 18.1%, compared to 11.8% in the case of an individual from the most advantaged social stratum (Q5) falling to Q1, which is 1.5 times lower. These results show that the father's socioeconomic status is decisive in the child's wages in the formal labor market when they are 20 to 27 years old.

Table 6 – Summary Statistics on Rank Disparities and Intergenerational Mobility for each sample

	Sample		
	(1)	(2)	(3)
	Sample 1	Sample 2	Sample 3
Median Children's Salary (R$)	1,763.77	1,557.08	1,490.5
Mean Parent Income Rank	48.7	50.1	49.8
P (Child in Q1 | Parent in Q1)	16.9%	18.1%	18.6%
P (Child in Q5 | Parent in Q1)	10.9%	14.8%	10.8%
P (Child in Q1 | Parent in Q5)	10.5%	11.8%	15.9%
P (Child in Q5 | Parent in Q5)	24.3%	27.5%	18.8%
Number of children	503,970	681,350	620,506

Notes: This table presents summary statistics on income and intergenerational mobility for each working sample: (i) sample 1 - Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS); (ii) Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS); (iii) Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS). The median salary consists of the median deflated December salary of the children in each sample. The deflation was calculated using the consumer price index (IPCA, in Portuguese) with the base year of 2018, so all monetary values are measured in 2018 Brazilian reals. The mean parent income rank is simply the calculation of the mean parent ranking for each sample. The rankings are calculated based on an open population, thus, children are assigned percentile ranks relative to all other young workers (aged 20-24, or 24-27, depending on the sample), not just the children in their respective samples. The same is true for parents, who are ranked relative to all other parents of students of each grade (9th grade or 12th grade, depending on the sample). Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively. We call the socioeconomic situation P (Child in Q1| Parent in Q1) as Poverty trap; the P (Child in Q5| Parent in Q1) as Upward mobility; P (Child in Q1| Parent in Q5) as Downward mobility and; the P (Child in Q5| Parent in Q5) as Remaining rich.

Blacks and Whites

In this section, we characterize the evolution of racial disparities in intergenerational mobility using the conceptual framework in Section 3, the rank-rank regression, and the transition matrices.

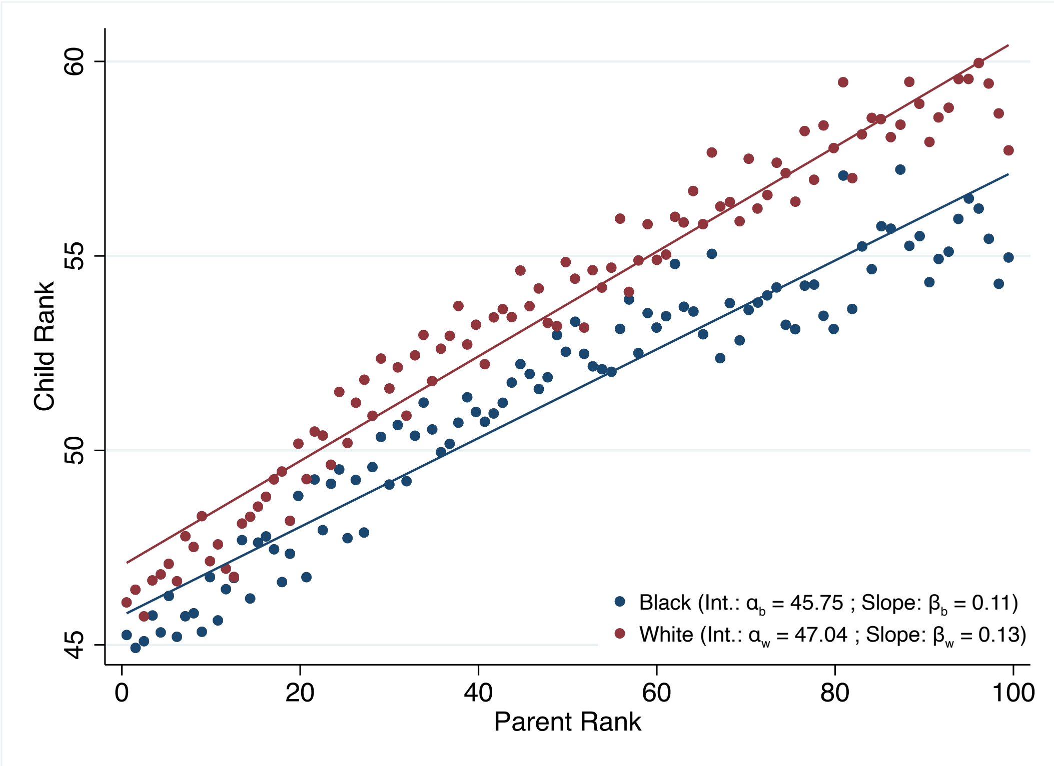
Figure 8 plots the mean salary rank of the children against the mean income rank of their parents for black and white children of Sample 1[8]. Therefore, for each racial group, the absolute and relative intergenerational mobility (α_r, β_r) was estimated based on specification 1.

For whites, the relative mobility estimated is $\beta_W = 0.13$, which means that a 10 percentile increase in parent's rank is associated with a 1.3 percentile increase in children's rank, on average. The intercept for whites is $\alpha_W = 47.0$, meaning that white children born in the lowest social stratum reach the 47th percentile on average. The interpretation of the coefficients for blacks follows the same logic. Therefore, blacks have lower rates of absolute mobility across the entire parental income distribution, but higher relative mobility. We tested whether differences in the slope and intercept coefficients across racial groups are statistically significant. The test results show that absolute and relative mobility statistically differs across racial groups (see Appendix B.2).

It is important to mention that the rank-rank regression and the scatter plot from Figure 8 quantify the degree of persistence for the whole distribution. Consequently, this statistic does not capture the status persistence for each group located on the higher social stratum or, the lower stratum, for instance, but the transition matrix does. Therefore, the transition matrices best analyze comparisons within the same distribution for different percentiles.

[8]We did the estimates for the three samples. However, since the results were very similar, we focused only on Sample 1. In Appendix C.1 and C.2, the reader can find the results for the other two samples.

Figure 8 – Empirical Estimates of Intergenerational Mobility and Racial Disparities for Sample 1



Notes: Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS).

Table 7 reports parts of the transition matrix of Sample 1, filtering the quintiles 1 and 5 of each generation. This transition matrix shows how much the parents' socioeconomic condition is associated with the children's salary in the formal labor market for whites and blacks when they are still young – in this case, between 24 and 27 years old[9]. On average, blacks come from a lower social stratum than whites. The mean parent income rank for blacks is 43.8, 8 percentiles below whites[10]. The mean child salary rank also differs. Black children's salary is three percentiles below whites, on average.

The results in Table 7 also show that the differences in mean ranks between

[9]Since we are still using Sample 1 to analyze the results here.
[10]This difference is higher than the absolute mobility shown in the Figure 2 since the distance between the curves increases along the parent ranking.

black and white children arise from the fact that blacks are more likely to fall from the top quintile (Q5) to the lower (Q1) (11.3% vs. 9.7%) and less likely to rise from the lower (Q1) to the top (Q5) (10.2% vs. 12.3%). That is, blacks are 1.6 percentage points more likely to experience downward mobility and 2.1 points less likely to experience upward mobility. Blacks are also more likely to stay in poverty (0.51 p.p. higher than whites) and less likely to remain in wealth (5.58 p.p. lower than whites).[11] Therefore, under the assumption that mobility rates remain constant across generations, these numbers show that the disparity between blacks and whites will evolve in the long-term.

Combining the results in Tables 6 and 7, it is possible to summarize some findings about intergenerational mobility in São Paulo. First, we see an apparent intergenerational persistence of wages, and the parent's socioeconomic status is decisive in the child's wages for the entire Sample 1. Additionally, a racial disadvantage exists in intergenerational mobility for blacks compared to whites. Black people have more difficulties moving up in social class or staying in wealth but have an easier time staying in poverty.

[11] The percentage point calculation is only based on the difference between the whites' and blacks' probabilities.

Table 7 – Summary Statistics on Rank Disparities and Intergenerational Mobility by Race for Sample 1

	White	**Black**
Mean Child Salary Rank	53.99	50.75
Mean Parent Income Rank	51.76	43.84
P (Child in Q1 | Parent in Q1)	16.62%	17.13%
P (Child in Q5 | Parent in Q1)	12.32%	10.24%
P (Child in Q1 | Parent in Q5)	9.77%	11.38%
P (Child in Q5 | Parent in Q5)	27.25%	21.67%
Number of children	*245,302*	*192,084*

Notes: This table presents summary statistics on income and intergenerational mobility for the working sample 1: Parents from children in the 12th grade (SARESP) and these same children when they were 24 to 27 years old in the formal labor market (RAIS). The mean child salary rank consists of the mean ranking for sample 1. The mean parent income rank calculates the mean parent ranking for sample 1. The rankings are calculated based on an open population, thus, children are assigned percentile ranks relative to all other young workers aged 24 to 27 years old, not only the children in their respective samples. The same is true for parents, who are ranked relative to all other parents of students in 12th grade. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.

Comparisons with other articles in the literature

The estimations found in this paper, presented in both Figure 7 and 8, point to differences in the same direction but with considerably distinct magnitudes from those shown in the literature. This section aims to explain the main reasons for the divergences between the parameters estimated here, with those found by Britto et al. (2022) in Brazil and those of Chetty et al. (2020) in the US.

In an analysis of the entire US population, Chetty et al. (2020) find that the relative mobility of the US is $\beta_{US} = 0.35$, and the absolute mobility is $\alpha_{US} = 32.5$. Furthermore, Britto et al. (2022) find that the relative mobility of Brazil is $\beta_{BRA} = 0.55$, and the absolute mobility is $\alpha_{BRA} = 35$. These differences are comparable since both studies use very similar techniques and data. The main reasons for these disparities are the higher intergenerational persistence and income inequality in

Brazil than in the US. In this study, however, we found relative mobility ranging from 0.06 to 0.13 and absolute mobility between 46 and 48 for our three samples (Figure 7).

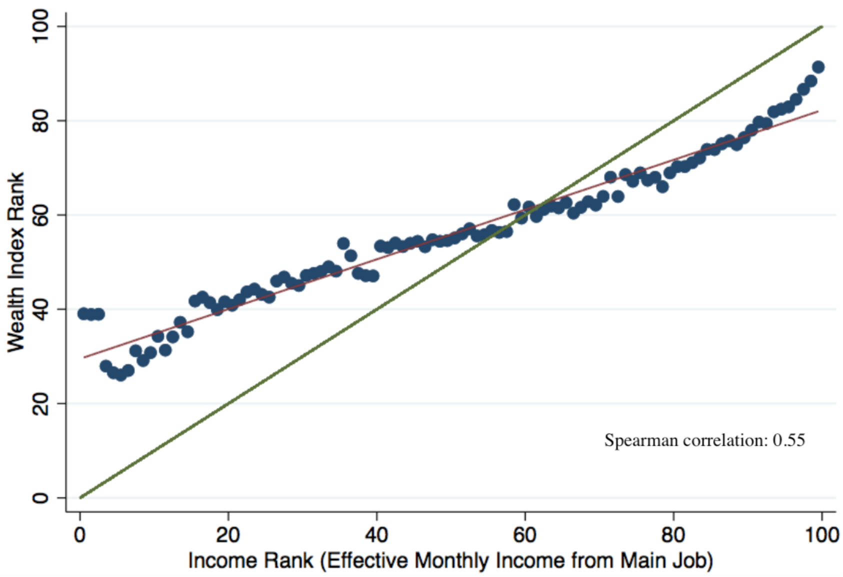
In the comparative race analysis, Chetty et al. (2020) find an average difference in absolute mobility between blacks and whites of 13.5 percentiles and Britto et al. (2022) of 7.1. In our study, the difference is around three percentiles. As mentioned in the introduction, although Chetty has inspired us, this study should not be considered a replication of his research in Brazil, as is the case with the article by Britto et al. (2022). Four main differences between our studies do not allow a direct comparison between the parameters.

First, our data is limited to the state of São Paulo and the formal labor market, so our samples do not represent all of Brazil's population. These first two differences are factors that overestimate our parameters for Brazilian mobility. The state of São Paulo overestimates since this is one of the most wealthy states of Brazil, and as shown by Britto et al. (2022), this state has higher upward mobility than the whole country. And concerning the formal labor market, in Section 4, we present an exercise with the PNAD Continuous (2016 to 2019), analyzing the differences in wage distribution between the formal and the informal market in Brazil (Figure 1). This figure reveals that there are more individuals with lower salaries in the informal market than in the formal, which corroborates the idea of overestimating the parameter mentioned. Britto et al. (2022) confirm this hypothesis in their paper by showing that the parameters point to greater absolute and relative mobility when limiting their analysis to the formal market.

Third, there is a difference concerning the measure of income we use. Chetty et al. (2020) analyze federal income tax returns data and use an income variable for parents and children to calculate the rankings. And the same occurs in the study by Britto et al. (2022), which has data from the Brazilian tax authority (Receita Federal do Brasil). However, the data used in this research paper was limited to assessing wages as children's income. Even though 75.3% of the average real monthly household income per capita is labor-based (IBGE (2022)), we have a measurement error in the income variable for not including the non-labor component (which includes dividends, rents, interest, and capital gains).

In order to better understand the relationship between wealth and wage measures, we did another exercise with PNAD Continuous (2016 to 2019) data. Figure 9 plots the income rank of individuals using the variable "Effective Monthly Salary from Main Job" against the Wealth indicator ranking. We created this indicator using PNAD's socioeconomic questions (similar to those used in the SARESP Wealth Index). We find that there is a correlation of 0.55 (Figure 9). Therefore, although we did not use the variable recommended by the literature because of data limitations, there is a strong correlation between the wage rank, the variable we are using, and the wealth rank.

Figure 9 – Correlation between salary rank and wealth index rank from PNAD (2016 to 2019)



Finally, there is also a limitation regarding the age we analyze the children already in the labor market. The literature understands that it is around age 35 that the peak of individuals' labor return occurs (Murphy & Welch (1990)), that age is when the return starts to reduce and become negative in the quadratic parameter. Chetty et al. (2020) can examine the children's income at the age of 31 to 37, but in this paper, we analyze them from 20 to 27 since this is the maximum age that we can find the individuals both in school (with the SARESP data) and in the labor market in the future (with the RAIS data). However, in Section 4 (under "Sample Definition"), we present an exercise showing the relationship between the wage ranking when RAIS workers are 20 to 27 years old compared to the wage ranking when they are 33 to 40 years old, and we find a correlation of 0.63, which is a reasonably high correlation.

There is a fifth difference between our studies, which derives from the above-mentioned differences, the ranking calculation. As mentioned, Chetty et al. (2020) and Britto et al. (2022) contain the entire universe of individuals from the US and Brazil. Therefore, the parent and children's rankings calculation is based on a closed population, the whole universe of individuals. However, the nature of our data precludes such a method because we are working with a subpopulation of individuals. Hence, our rankings are calculated based on an open population, that is, people from our working samples are ranked relative to *all* individuals of RAIS from the same age and *all* individuals of SARESP from the same school grade, and not only within the subpopulation of that respective working sample (see Section 4, "Variable Definitions" for more details of the ranking construction). We chose this method to determine better where people from our sub-sample are compared to their equivalents in a larger population. Doing the ranking only on the smaller sample would not give us a number put into perspective of the entire comparable population. This difference in how the ranking is calculated makes it impossible to compare specifically the relative mobility parameter between our study and others because it changes how we compare individuals relatively.

Highlighted the main differences between our study and recent literature, this research paper is not a replication of Chetty et al. (2020) and Britto et al. (2022), but rather a complement, in which we use different data, with a greater level of detail in terms of school characteristics, to bring to the research agenda new discussions related to how education and different school characteristics can influence intergenerational mobility for different racial groups.

School and Educational characteristics

In this section, we will explore what this research paper brings new to the literature: to understand how school and educational characteristics that vary across black and white students are associated with their future earnings and how these factors seem to affect the intergenerational racial gaps. We continue to use Sample 1 to illustrate all effects. The results of the other samples (2 and 3) will be discussed only in cases where the conclusions differ from Sample 1. If the results are similar, there will be no discussion in the text, but all results will be exposed in Appendix B.

School Performance

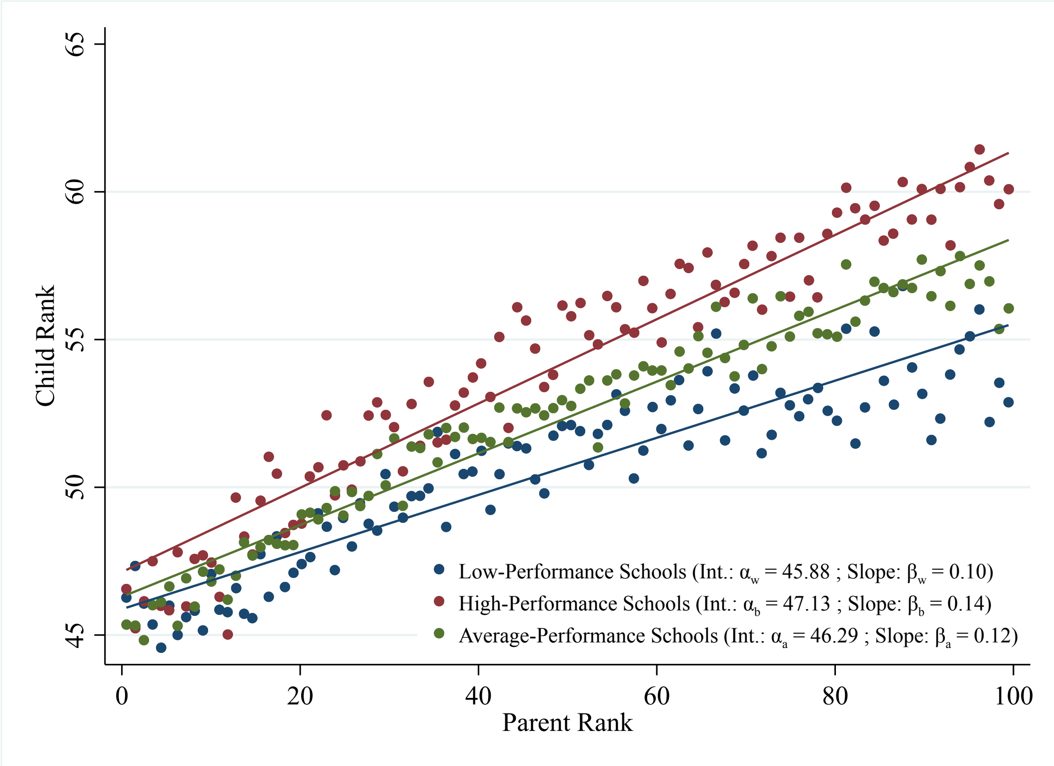
We start by analyzing how school performance in a high-stake exam is associated with their student's future earnings and how it affects the racial gap between whites and blacks. We measure school performance according to a school average of Portuguese and Mathematics scores that the 12th-grade students took in the SARESP exam without the grade of the respective student (see more details in Section 4, "Variable Definition"). In doing so, we are creating "synthetic" schools, in which each different grade represents another school. We estimate the school's performance according to the pool of students attending that school, excluding the individual's respective test scores. We use this measure to separate the effect of student performance from school performance. First, we document the differences in intergenerational mobility between the three school categories (high-performance, average-performance, and low-performance schools). Then, we run the transition matrices to see the differences in upward and downward mobility for students by race and by each school's category.

The scores are ranked, divided into quintiles, and classified among the three possible categories. Sample 1 contains 284,378 "synthetic" schools, of which 70,599 are in the "High-Performance Schools" category (quintile 5), 61,198 in the "Low-Performance Schools" category (quintile 1), and the rest (152,581) are in the "Average-Performance Schools" category.

Figure 10 shows that the three categories of schools have relative and absolute mobility distinct from each other. The high-performance schools in SARESP have a higher absolute mobility rate than the other two categories. However, this rate is not uniformly distributed, the difference is increasing across the parental income rank. For example, 12th-grade students from the high-performance schools born to parents at the 25th percentile of the income distribution reach the 50th percentile on average, 2 percentiles above students from the low-performance schools with parents in the same percentile rank. Nonetheless, students from the high-performance schools with parents at the 90th percentile reach the 60th percentile on average, 5 percentiles above youth from the low-performance schools with parents at the identical rank distribution.

This finding highlights that the education effect differs depending on the parent's income rank. More specifically, the positive effect of a school with high scores is increasing along with the parent's ranking. This may indicate that other mechanisms trap the individual in poverty or cause them to persist in poverty for generations, i.e., a high-performance school alone does not seem to be a sufficient mechanism to lift the individual out of poverty, or at least to end the relative differences. For instance, Durlauf et al. (2006) argue that many conditions can keep an individual or a group of individuals in poverty, among them dysfunctional institutions and neighborhood effects. They exemplify neighborhood effects such as networks, role models, and aspirations that can create hard-to-escape pockets of poverty. Similar individuals in different socioeconomic environments develop different preferences and beliefs that can transmit poverty or affluence from generation to generation. Kraay & McKenzie (2014) discuss other mechanisms, from macro and micro perspectives, such as savings and coordination failure, nutrition and lumpy investments, and behavioral and geographic poverty traps. In other words, the literature shows us that the persistence of generations in poverty has several causes, and education alone may be unable to get individuals out of this scenario.

Figure 10 – Empirical Estimates of Intergenerational Mobility for Different School Performance Categories for Sample 1



Notes: Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS).

Table 8 presents the upward and downward mobility rates for high- and low-performance school students. For example, among children with parents in the bottom quintile, 13% of youth from high-performance schools rise to the top quintile, but only 10% of students from low-performance schools, representing a difference of 23%. On the other hand, among children with parents in the top quintile and attending high-performance schools, on average, 30% remain in this quintile, but only 19% of the low-performance do, revealing a difference of 35%. At the same time, students from low-score schools have 16% more chances of falling from the top to the bottom quintile than students from high-score schools. If we again assume that mobility rates remain constant across generations, these numbers show that the disparity between students from low and high-score schools will evolve across generations. This finding highlights that low-performance schools are detrimental to the current generation and can also affect future generations' income and social mobility.

Table 8 – Intergenerational Mobility by High and Low-Performance Schools in SARESP for Sample 1

	High-Performance	Low-Performance
P (Child in Q1 | Parent in Q1)	16.50%	16.95%
P (Child in Q5 | Parent in Q1)	12.56%	9.92%
P (Child in Q1 | Parent in Q5)	9.50%	11.36%
P (Child in Q5 | Parent in Q5)	28.90%	19.14%
Number of children	*97,578*	*82,705*

Notes: This table presents the results of the extremes of the transition matrix for students from the high and low-performance schools in SARESP for Sample 1. Sample 1: Parents from children in the 12th grade (SARESP) and these same children when they were 24 to 27 years old in the formal labor market (RAIS). School performance is measured according to a school average of Portuguese and Mathematics scores that the 12th-grade students took in the SARESP exam without the grade of the respective student. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.

The second part of this subsection intends to answer how school performance in a high-stake exam affects the racial gap between whites and blacks. Table 9 presents the four extremes of the transition matrix for students from high and low-test scores schools by race. In column 3 we calculated the racial gap, which is the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the analysis of the racial gap indicator in column 3 (see Section 4, "Variable Definition" for all the different interpretations of this indicator). We found that high-performance schools have higher upward mobility for whites and blacks and lower downward mobility for both. Regarding the racial gap, the high-test scores schools are more favorable to blacks than low-performance schools in the cases of upward mobility and remaining rich, the two most economically favorable scenarios, but this is not verified for downward mobility and poverty trap scenarios. In relative terms, high-performance schools do not greatly affect blacks' persistence in poverty or their chances of falling socially. However it favors blacks for upward mobility and remaining in wealth.

In general terms, Figure 10 and Table 9 together reveal that, regardless of race, students from top schools have higher absolute mobility than other schools. Nevertheless, there is no significant change in relative mobility, suggesting that students from high-performance schools still depend significantly on their parents' position. Furthermore, students from this same school category are more likely to rise socially and less likely to fall socially compared to students from low-performance schools.

Moreover, regarding racial disparities, we see a decrease in the racial gap in high-performance schools in economically favorable situations (such as upward mobility and staying in wealth) but not in situations of disadvantage (such as downward mobility and poverty trap). This last finding corroborates, to some extent, with Arias et al. (2004), who show that equalizing access to quality education is one of the critical factors in reducing racial income inequality in Brazil.

Table 9 – Intergenerational Mobility by High- and Low-Performance Schools in SARESP for Sample 1

	(1) White	(2) Black	(3) = (1)/(2) Racial Gap	(4) Interpretation of (3)
Panel A. Poverty Trap: P(Q1|Q1)				
High-Performance	15.60%	17.27%	0.90	Low-Performance schools are
Low-Performance	16.00%	17.32%	0.92	more favorable for blacks
Panel B. Upward Mobility: P(Q5|Q1)				
High-Performance	13.66%	11.71%	1.17	High-Performance schools are
Low-Performance	11.93%	9.29%	1.28	more favorable for blacks
Panel C. Downward Mobility: P(Q1|Q5)				
High-Performance	8.89%	10.40%	0.85	Low-Performance schools are
Low-Performance	10.58%	11.61%	0.91	more favorable for blacks
Panel D. Remaining Rich: P(Q5|Q5)				
High-Performance	30.68%	27.15%	1.13	High-Performance schools are
Low-Performance	22.28%	17.70%	1.26	more favorable for blacks
Number of observations	*White*	*Black*		
High-Performance	*57,700*	*28,670*		
Low-Performance	*30,995*	*40,238*		

*Notes: This table presents the results of the extremes of the transition matrix for students from the high- and low-performance schools in SARESP by race, for Sample 1. Sample 1: Parents from children in the 12th grade (SARESP) and these same children when they were 24 to 27 years old in the formal labor market (RAIS). School performance is measured according to a school average of Portuguese and Mathematics scores that the 12th-grade students took in the SARESP exam without the grade of the respective student. The racial gap is calculated as the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the interpretation of the racial gap indicator in column 3. This indicator measures how much whites' probability of being in the respective socioeconomic situation (poverty trap, upward or downward mobility, or remaining rich) is higher (> 1) or lower (< 1) than blacks'. This indicator should be interpreted with caution because the interpretation changes in each socioeconomic situation, so we have already added column (4) with the conclusion of its interpretation. We consider better the educational setting that is more favorable for the blacks and therefore is going in the direction of reducing racial differences. The interpretation for each row should be made as follows: (i) Poverty trap situation (P (Child in Q1| Parent in Q1)): the **higher** the racial gap, more favorable for blacks; (ii) Upward mobility (P (Child in Q5| Parent in Q1)): the **lower** the racial gap, more favorable for blacks; (iii) Downward mobility (P (Child in Q1| Parent in Q5)): the **higher** the racial gap, more favorable for blacks; (iv) Remaining rich (P (Child in Q5| Parent in Q5)): the **lower** the racial gap, more favorable for blacks. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.*

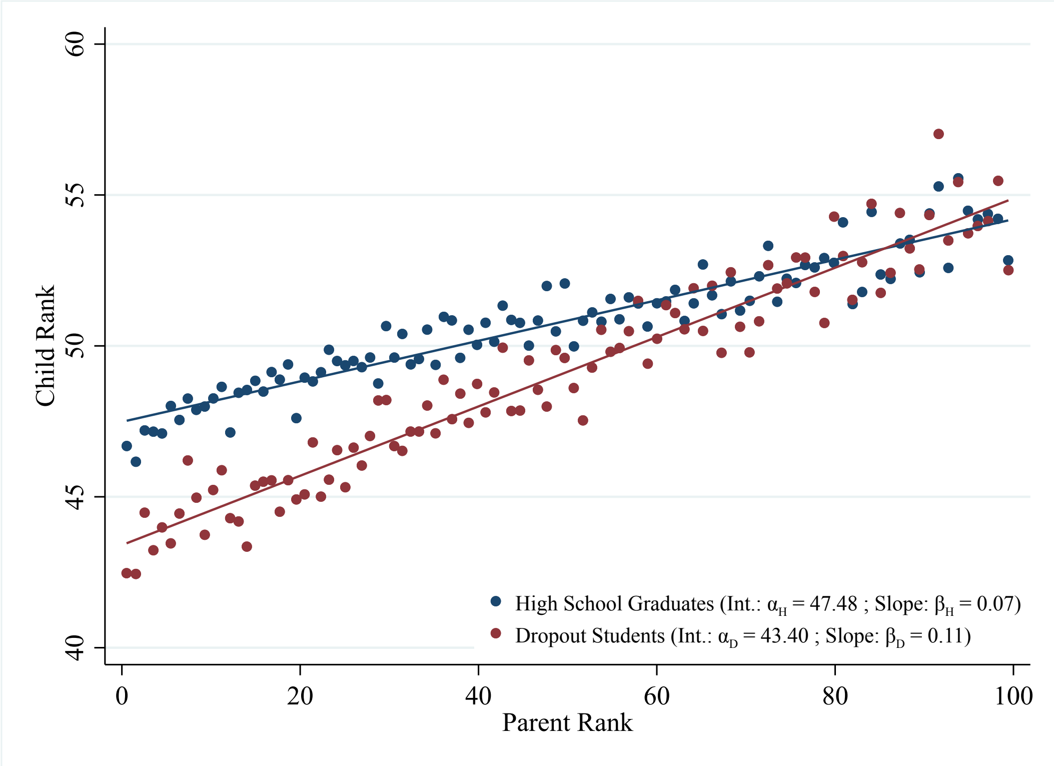
Educational Trajectory

In this subsection, we will analyze the effects of educational trajectory on children's future earnings and racial gap. Since the samples we work with are either 12th-grade or 9th-grade, to track the student's trajectory, we will use Sample 3, which refers to ninth-grade students. Our school trajectory variable is divided into two categories: (i) students who dropped out school; and (ii) high school graduates (for more details see Section 4, subsection "Variable Definition").

Figure 11 presents the absolute and relative mobilities for the different educational paths. Note that individuals who have completed high school have higher absolute and relative mobility. The absolute mobility of high school graduates is $\alpha_H = 47.5$, which is, on average, 9.4% higher than the absolute mobility of students who dropped out of school. These differences show that education seems to increase the expected income rank of children born to below-median-income parents. In addition, high school graduates also have higher relative mobility, suggesting that education makes the parent rank less deterministic in the child's arrival quintile.

Table 10 presents some results that complement those in Figure 11. For instance, it seems that one of the reasons for the greater absolute mobility of high school graduates is the higher probability of upward mobility. That is, a high school diploma positively affects upward mobility rates. First, analyzing only whites, take the ratio between the probability of students with a high school diploma experiencing upward mobility (13.29) over the probability of dropout students in the same situation (9.28), and we arrive at a relative difference of 1.43. This number basically tell us that children with a high school diploma and parents in the bottom quintile are 43% more likely to move up to the top quintile than students who dropped out of school. Doing the same calculations for blacks, we also found a difference of 44%. In addition, students who have dropped out of school also have higher chances of downward mobility.

Figure 11 – Empirical Estimates of Intergenerational Mobility for Different Educational Trajectories for Sample 3



Notes: Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

We follow this subsection to analyze the effects of different educational paths on the racial gap. In other words, we want to understand how the racial gap varies among the two trajectories presented above. To do so, we have columns (3) and (4) of Table 10. In column (3), we calculated the racial gap, which is the ratio between the white's probability (column 1) and the black's probability (column 2). And in column (4), we put the analysis of the racial gap indicator in column 3 (see Section 4, "Variable Definition" for all the different interpretations of this indicator).

Looking at column (4) of Table 10, we see that the racial gap is smaller among individuals with a high school diploma, for upward mobility and remaining rich scenarios. However, it does not seem to decrease racial differences in downward mobility and poverty trap cases. This corroborates Chetty et al. (2020), who show that high school dropout rates cannot directly explain the sharp disparities observed in outcomes, between blacks and whites, at younger ages.

Table 10 – Intergenerational Mobility by Race for Different Educational Trajectories for Sample 3

	(1) White	(2) Black	(3) = (1)/(2) Racial Gap	(4) Interpretation of (3)
Panel A. Poverty Trap: P(Q1|Q1)				
High School Graduates	16.75%	17.78%	0.94	Dropout students
Dropout Students	19.20%	19.45%	0.98	are more favorable for blacks
Panel B. Upward Mobility: P(Q5|Q1)				
High School Graduates	13.29%	11.87%	1.12	Both educational situations are
Dropout Students	9.28%	8.27%	1.12	equally favorable for blacks
Panel C. Downward Mobility: P(Q1|Q5)				
High School Graduates	13.55%	15.80%	0.86	Dropout students
Dropout Students	14.21%	15.43%	0.92	are more favorable for blacks
Panel D. Remaining Rich: P(Q5|Q5)				
High School Graduates	21.89%	18.62%	1.18	High School Graduates
Dropout Students	24.46%	18.35%	1.33	are more favorable for blacks
Number of observations	Whites	Blacks		
High School Graduates	*184,772*	*178,314*		
Dropout Students	*49,770*	*49,663*		

Notes: This table presents the results of the extremes of the transition matrix for students from different educational trajectories by race, for Sample 3. Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS). Each of the three categories of this variable was constructed as follows: (1) Students who dropped out in 9th grade - in this category, we only include students who were: (i) last found in the enrollment database in some year before 2014; and (ii) were enrolled in 9th grade that year; (2) Students who dropped out either in 10th or 11th grade - in this category, we only include students who were: (i) last found in the enrollment database in some year before 2014; and (ii) were enrolled in 10th or 11th grade that year; (3) Students found in 12th grade - in this category, we included only students found in 12th grade in any year of the enrollment base. The racial gap is calculated as the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the interpretation of the racial gap indicator in column 3. This indicator measures how much whites' probability of being in the respective socioeconomic situation (poverty trap, upward or downward mobility, or remaining rich) is higher (> 1) or lower (< 1) than blacks'. This indicator should be interpreted with caution because the interpretation changes in each socioeconomic situation, so we have already added column (4) with the conclusion of its interpretation. We consider better the educational setting that is more favorable for the blacks and therefore is going in the direction of reducing racial differences. The interpretation for each row should be made as follows: (i) Poverty trap situation (P (Child in Q1| Parent in Q1)): the ***higher*** *the racial gap, more favorable for blacks; (ii) Upward mobility (P (Child in Q5| Parent in Q1)): the* ***lower*** *the racial gap, more favorable for blacks; (iii) Downward mobility (P (Child in Q1| Parent in Q5)): the* ***higher*** *the racial gap, more favorable for blacks; (iv) Remaining rich (P (Child in Q5| Parent in Q5)): the* ***lower*** *the racial gap, more favorable for blacks. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.*

Social Segregation in School

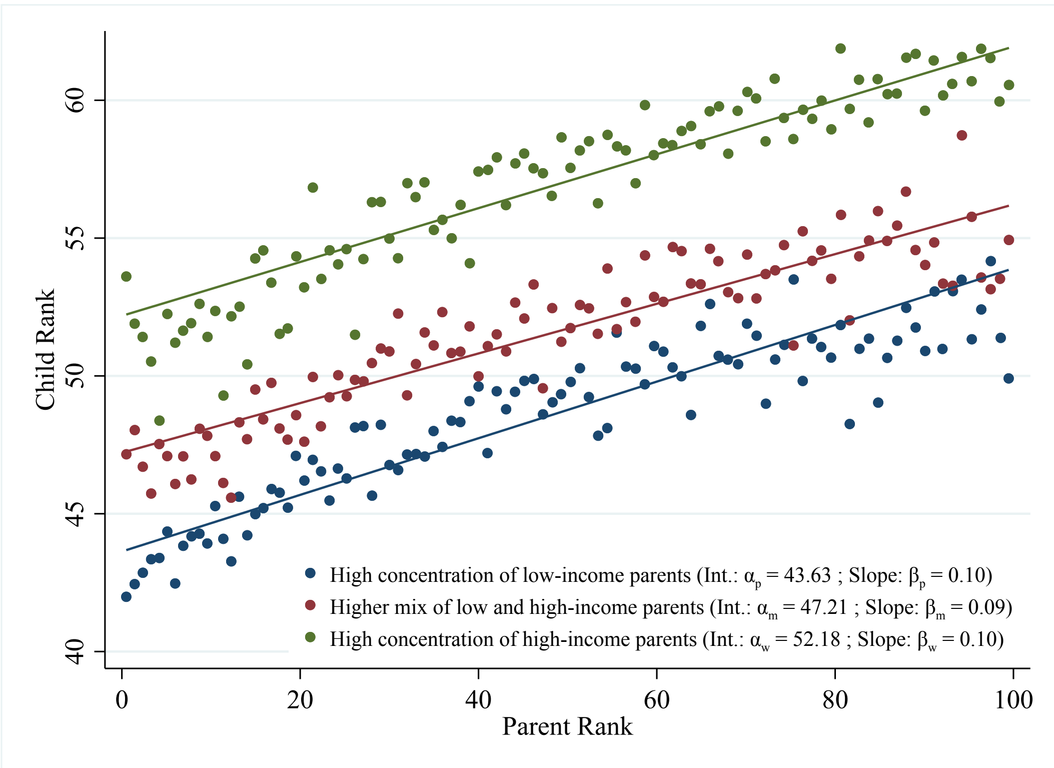
In this subsection, we analyze whether more social diversity in schools can affect the children's future earnings and the racial gap between whites and blacks in intergenerational mobility. To do so, we start by estimating the rank-rank regression for each school's category (based on levels of social inequality). And we follow with the transition matrices, showing the intergenerational mobility for the different types of schools (in terms of social inequality) by race.

Figure 12 shows the intergenerational mobility estimates for students who attended schools of different categories in terms of social inequality:(i) a high concentration of high-income parents; (ii) a high concentration of low-income parents; and (iii) a higher diversity between high and low-income parents (see Section 4, "Variable Definition" for more details on how this variable is constructed). Students who attended schools with a higher concentration of wealthy parents in the 12th grade have absolute mobility of $\alpha_W = 52.2$, i.e., students who study in this type of school born to the lowest-income parents reach the 52.2 percentile on average. The relative mobility (slope) is $\beta_W = 0.09$, i.e., a 10 percent increase in the parents' ranking is associated with an average 0.9 percent increase in the children's ranking.

The blue curve represents students who attended schools with a higher concentration of low-income families in the 12th grade. These groups have comparable relative mobility to the other groups ($\beta_P = 0.10$) but have uniformly lower rates of absolute mobility across the parental income distribution. In this case, the intercept is $\alpha_P = 43.63$, which is almost 20% lower than students from the same socioeconomic level but who attended schools with a higher concentration of high-income families.

Finally, for students who studied at schools with a more significant socioeconomic variation (i.e., a similar amount of high- and low-income families) in the 12th grade, the relative mobility is also comparable to the other two groups ($\beta_M = 0.09$). Absolute mobility, meanwhile, is $\alpha_M = 47.2$, 10.5% lower than students from schools with more high-income families and 8.5% higher than students from schools with more low-income families. This first finding suggests that the whole population benefits more when they are among wealthier peers. Comparing the absolute mobility rates, we conclude that individuals from the lowest socioeconomic strata in schools with more affluent families have a higher expected future income rank compared to individuals from the same social strata in schools with more low-income families.

Figure 12 – Empirical Estimates of Intergenerational Mobility for Different School Categories, in terms of social inequality for Sample 1



Notes: Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS).

We continue our investigation by asking how this benefit of attending schools with more peers from high-income families is distributed among racial groups. From this perspective, two crucial questions are: "Are schools with a higher concentration of wealthy parents, besides benefiting everyone, benefiting each race individually? And is it able to reduce the racial gap?".

The short answer to the first question is “Yes”. Table 11 shows that whites and blacks who attend schools with a higher concentration of wealthy parents have: higher probabilities of upward mobility and of remaining rich than in the other schools; and lower probabilities of downward mobility and poverty trap compared to the other two school’s categories. Thus, the result of the table corroborates the one presented in the figure above for both races. This finding goes in line with Chetty et al. (2022), who showed that communities in which people with low SES interact less with people with high SES exhibit less upward income mobility across generations. Moreover, the degree of friendship between people with low and high SES is strongly associated with upward income mobility.

Concerning the second question, we need to analyze columns (3) and (4) of Table 11, which follow the same reasoning as Tables 9 and 10. Looking at column (4) of Table 11, we see that schools with a high concentration of low-income families have the slightest racial gap in all socioeconomic situations. Although whites and blacks benefit more in absolute terms in schools with a higher concentration of wealthy parents, whites benefit more in relative terms, increasing the racial gap. Showing this argument with the numbers in the table, let’s start by looking only at students in schools with a higher concentration of high-income parents. If the reader takes the ratio between the probability of whites experiencing upward mobility over the probability of blacks in the same situation, we arrive at a relative difference of 1.26. This number shows that whites are 26% more likely than blacks to experience upward mobility in this school category. However, if we do the same calculations in upward mobility for students attending schools with a high concentration of low-income families or more diverse schools, we find that whites are 12% more likely to experience upward mobility than blacks, proving the previous argument.

In the socioeconomic situation “Remaining Rich”, note that white students from schools with a high concentration of wealthy parents also benefit more than blacks compared to the other school categories. That is, it has a more significant racial gap. In the case of downward mobility and poverty trap, although students in this same school type are less likely to experience these situations, the difference between blacks and whites is also greater than in schools with a higher concentration of low-income parents.

From the analyses presented above, school policies related to increasing socioeconomic diversity will have different outcomes. On the one hand, policies to increase wealthy peers in schools may benefit all groups positively. However, on the other, this will also potentiate relative differences between racial groups.

One discussion that emerges from these results is why schools with a high concentration of high-income parents generate greater intergenerational mobility than other school categories. This debate relates to the existing literature on economic connectivity. For example, Loury (1977) and Lin & Dumin (1986) show that economic mobility can be facilitated by connections with people who can provide access to information and employment opportunities or with people who can help shape aspirations.

Table 11 – Intergenerational Mobility by Race for different school categories in terms of social concentration for Sample 1

	(1) White	(2) Black	(3) = (1)/(2) Racial Gap	(4) Interpretation of (3)
Panel A. Poverty Trap: P(Q1|Q1)				
Concentration of low-income families	20.14%	19.41%	1.04	Schools with a high concentration
Concentration of high-income families	12.36%	13.38%	0.92	of low-income families are
Higher diversity (low and high-income families)	14.52%	16.48%	0.88	more favorable for blacks
Panel B. Upward Mobility: P(Q5|Q1)				
Concentration of low-income families	9.72%	8.65%	1.12	Schools with a high concentration
Concentration of high-income families	19.26%	15.31%	1.26	of low-income families and with a
Higher diversity (low and high-income families)	11.98%	10.68%	1.12	higher diversity are more favorable for blacks
Panel C. Downward Mobility: P(Q1|Q5)				
Concentration of low-income families	14.91%	13.68%	1.09	Schools with a high concentration
Concentration of high-income families	8.10%	9.66%	0.84	of low-income families are
Higher diversity (low and high-income families)	10.71%	12.00%	0.89	more favorable for blacks
Panel D. Remaining Rich: P(Q5|Q5)				
Concentration of low-income families	18.23%	15.62%	1.17	Schools with a high concentration
Concentration of high-income families	33.04%	26.88%	1.23	of low-income families are
Higher diversity (low and high-income families)	23.24%	19.47%	1.19	more favorable for blacks
Number of observations	*White*	*Black*		
Concentration of low-income families	*45,453*	*42,779*		
Concentration of high-income families	*57,114*	*30,590*		
Higher diversity (low and high-income families)	*46,779*	*40,179*		

*Notes: This table presents the results of the extremes of the transition matrix for students from types of school in terms of social concentration, for Sample 1. Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS). The construction of this categorical variable was based on another variable that measures social inequality in schools. The measure used was created at the school level, such that: Social Inequality in Schools = $\frac{X}{Y}$, where X = Number of students from the wealthiest quintile of the total population; and Y = Number of students from the lowest quintile of the total population. The total population here is the all students from 12th grade. So, per school, we counted the number of students that belong to the highest quintile and the lowest quintile based on the parent's Wealth Index. Each of the three school categories was constructed as follows: (i) Schools with a high concentration of high-income parents - represent the last quintile of the social inequality variable (top 20%); (ii) Schools with a high concentration of low-income parents - represent the first quintile of the social inequality variable (bottom 20%); (iii) Schools with a higher mix of low and high-income parents - represent the third quintile of the social inequality variable since it is in this quintile that the variable crosses the 1 (same number of high and low-income families in the school). The racial gap is calculated as the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the interpretation of the racial gap indicator in column 3. This indicator measures how much whites' probability of being in the respective socioeconomic situation (poverty trap, upward or downward mobility, or remaining rich) is higher (> 1) or lower (< 1) than blacks'. This indicator should be interpreted with caution because the interpretation changes in each socioeconomic situation, so we have already added column (4) with the conclusion of its interpretation. We consider better the educational setting that is more favorable for the blacks and therefore is going in the direction of reducing racial differences. The interpretation for each row should be made as follows: (i) Poverty trap situation (P (Child in Q1| Parent in Q1)): the **higher** the racial gap, more favorable for blacks; (ii) Upward mobility (P (Child in Q5| Parent in Q1)): the **lower** the racial gap, more favorable for blacks; (iii) Downward mobility (P (Child in Q1| Parent in Q5)): the **higher** the racial gap, more favorable for blacks; (iv) Remaining rich (P (Child in Q5| Parent in Q5)): the **lower** the racial gap, more favorable for blacks. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.*

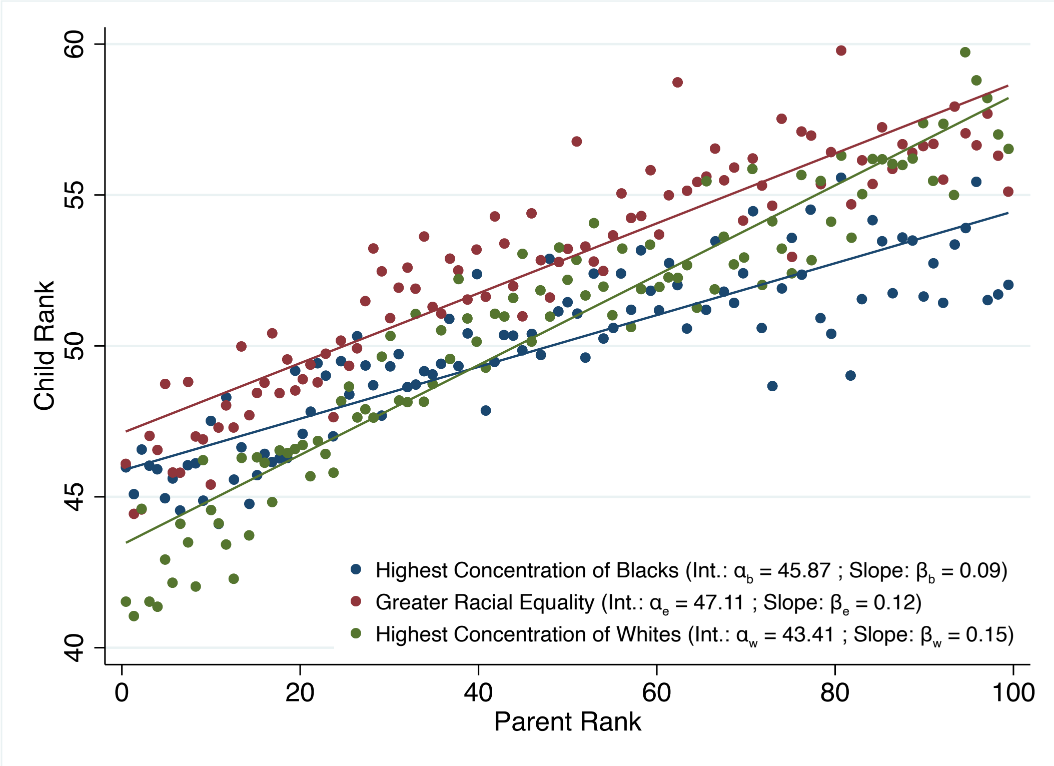
Racial Segregation in School

In this subsection, we follow the same logic as the previous one, but now for racial segregation in schools.

Figure 13 presents intergenerational mobility for students in schools with different racial concentrations: (i) high concentration of whites; (ii) high concentration of blacks; and (iii) greater racial equality (for more details see Section 4, "Variable Definition"). In general, this figure shows that schools with more racial diversity are associated with greater intergenerational mobility. Schools with more racial equality have a higher absolute mobility rate than other schools' categories throughout the distribution and higher relative mobility compared to schools with more whites. That is, schools with a higher concentration of whites are worse in terms of both absolute and relative mobility and therefore represent the type of school where the parent's economic position matters most for the children's future earnings. On the other hand, relative mobility in schools with more blacks is higher than in schools with greater racial equality. This means that where there is a higher concentration of blacks, economic capital has less value in intergenerational mobility; the parents' economic position makes less difference.

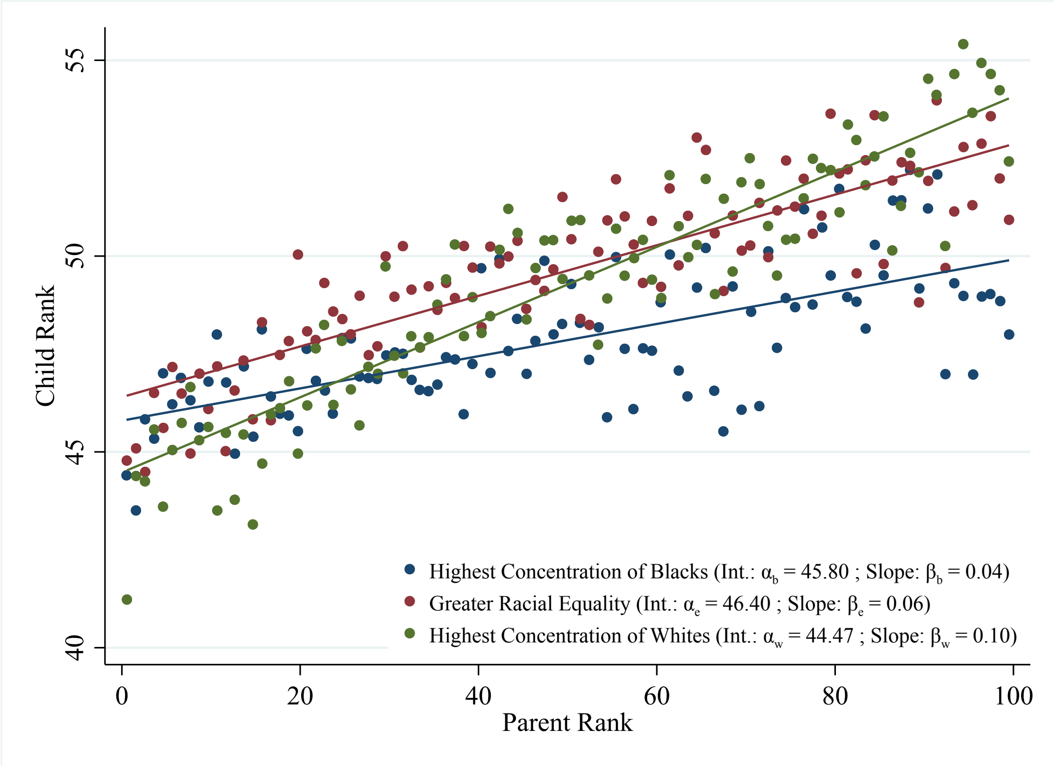
Figure 14 presents the results for Sample 3, which refers to 9th-grade students. This subsection is the only one we present results from another sample because they differed slightly from those found in Sample 1. In this case, as we are referring to 9th-grade students, we found lower relative mobility in all subgroups, which is already expected (see Subsection "Full Sample" in Section 5). However, in relative terms, the parent rank is even more deterministic for the child's arrival quintile in schools with a higher concentration of white students than the other two categories. That is, our findings show that for 9th grade, schools with a higher proportion of white students benefit children from parents in the upper quintiles of the distribution more than schools with the same proportion of white and black students, unlike the case for 12th grade.

Figure 13 – Empirical Estimates of Intergenerational Mobility for Different School Categories, in terms of racial concentration for Sample 1



Notes: Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS).

Figure 14 – Empirical Estimates of Intergenerational Mobility for Different School Categories, in terms of racial concentration for Sample 3



Notes: Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

As with the other sections, our next questions are concerning the heterogeneity analysis by race, and what happens to racial gaps in schools of different racial segregation categories. Table 12 shows the four ends of the transition matrix of Sample 1 by race and type of school. First, looking at intergenerational mobility for blacks, we see that schools with the same proportion of white and black students have higher upward mobility rates than the other two school categories and a lower probability of staying in poverty. Moreover, this same category, along with schools with a higher concentration of white students, is the most likely to keep black students in wealth. However, for some reason that has yet to be better studied in the literature, the transition matrix also shows that schools with the same proportion of white and black students have higher downward mobility rates than schools in the other two categories.

For white students, schools with the same proportion of white and black students provide higher upward mobility and staying-in-wealth rates, as well as lower downward mobility rates, when compared to schools in the other two categories of racial concentration.

Still analyzing the categories of schools separately, it is also worth noting that schools with higher concentrations of white students have more significant discrepancies in the probabilities of each socioeconomic situation, having very high chances of keeping students in poverty or wealth. On the other hand, schools with a higher concentration of black students have the worst rates in terms of downward mobility (higher probabilities) and remaining rich (lower probabilities). But this category provides higher rates of upward mobility than schools with a higher concentration of white students.

Analyzing now the racial gap, we need to look at columns (3) and (4) of Table 12, which follow the same reasoning as Tables 9-11. Looking at column (4) of Table 12, we see that schools with a higher proportion of white students are associated with smaller racial gaps in almost all the socioeconomic scenarios presented (for poverty trap, downward mobility, and remaining rich). In addition, schools with greater racial diversity are also associated with smaller racial gaps in two scenarios: the poverty trap and upward mobility. Nevertheless, this same school category is not the best at preventing blacks from falling in social class or keeping them in wealth. These results suggest that racial integration in schools alone seems insufficient to close the existing racial gap.

In the literature, there is an extensive discussion about racial segregation in neighborhoods, but little research has been done regarding racial segregation in schools, specifically. In the case of the State of São Paulo, the dominant rule regarding where students will be allocated is the house's proximity to the school. Therefore, we understand that racial segregation in schools can be considered, in this case, a good proxy for neighborhood racial segregation. Therefore, next, we explore the relationship between our results and this literature.

The impacts of residential segregation have been studied in the literature for some time. One well-known explanation for the impacts of residential segregation on racial gaps is from Douglas & Massey (1993). They find that blacks and whites can have different outcomes because they tend to live in different neighborhoods. More specifically, they show that poverty's effects will be more harmful in black neighborhoods than in whites within a racially segregated city. This result is related to the results in table 8. Examining the Poverty Trap line, we see that blacks are the most disadvantaged in both cases of more racially segregated schools. Taking the ratio between the blacks over the whites' probability, we find that blacks are 15% more likely than whites to be in poverty in schools with a higher concentration of blacks and 2% more likely in schools with a higher concentration of whites.

Douglas & Massey (1993) also show through simulations that when the residential environment is racially integrated, increased poverty is experienced by blacks and whites equally. Our results also relate to this finding because, looking at the poverty line, the racial gap for schools with a similar proportion of blacks and whites is 0.98, which means that blacks' and whites' probabilities of experiencing poverty traps are almost the same, with only a 2% difference. In this same literature, Chetty et al. (2020) explore the effects of racial segregation in neighborhoods on intergenerational mobility and find that only reducing residential segregation is insufficient to close the black-white gap. This result corroborates our results that racial integration in schools alone seems insufficient to close the existing racial gap.

Table 12 – Intergenerational Mobility by Race for different school categories in terms of racial concentration for Sample 1

	(1) White	(2) Black	(3) = (1)/(2) Racial Gap	(4) Interpretation of (3)
Panel A. Poverty Trap: P(Q1|Q1)				
High concentration of whites	20.68%	21.11%	0.98	Schools with a high concentration
Greater racial equality	15.91%	16.25%	0.98	of whites and greater racial equality
High concentration of blacks	15.30%	17.55%	0.87	are more favorable for blacks
Panel B. Upward Mobility: P(Q5|Q1)				
High concentration of whites	10.70%	7.80%	1.37	Schools with a greater racial
Greater racial equality	12.72%	11.63%	1.09	equality are more favorable
High concentration of blacks	11.85%	9.54%	1.24	for blacks
Panel C. Downward Mobility: P(Q1|Q5)				
High concentration of whites	10.30%	9.96%	1.03	Schools with a high concentration
Greater racial equality	9.48%	12.91%	0.73	of whites are more favorable
High concentration of blacks	11.67%	12.31%	0.95	for blacks
Panel D. Remaining Rich: P(Q5|Q5)				
High concentration of whites	25.88%	22.05%	1.17	Schools with a high concentration
Greater racial equality	26.85%	21.92%	1.22	of whites are more favorable
High concentration of blacks	20.88%	16.94%	1.23	for blacks
Number of observations	White	Black		
High concentration of whites	*34,378*	*10,046*		
Greater racial equality	*21,655*	*21,948*		
High concentration of blacks	*15,014*	*27,657*		

Notes: This table presents the results of the extremes of the transition matrix for students from types of school in terms of racial concentration by race, for Sample 1. Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS). The construction of this categorical variable was based on another variable that measures racial inequality in schools. The measure used was created at the school level, such that: Racial Inequality in Schools = $\frac{X}{Y}$*, where X = Number of white students in 12th grade and the respective school; and Y = number of black students in 12th grade and the respective school.Next, we created the categorical variable, racial concentration in schools, which has three categories: (i) Schools with the highest concentration of whites - represents the highest decile of the social inequality variable (top 10%); (ii) Schools with the highest concentration of blacks - represents the lowest decile of the social inequality variable (bottom 10%); (iii) Schools with greater racial equality - represents the fourth decile of the social inequality variable since it is in this decile that the variable crosses the 1 (same proportion of whites and blacks in school). The racial gap is calculated as the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the interpretation of the racial gap indicator in column 3. This indicator measures how much whites' probability of being in the respective socioeconomic situation (poverty trap, upward or downward mobility, or remaining rich) is higher (> 1) or lower (< 1) than blacks'. This indicator should be interpreted with caution because the interpretation changes in each socioeconomic situation, so we have already added column (4) with the conclusion of its interpretation. We consider better the educational setting that is more favorable for the blacks and therefore is going in the direction of reducing racial differences. The interpretation for each row should be made as follows: (i) Poverty trap situation (P (Child in Q1| Parent in Q1)): the* ***higher*** *the racial gap, more favorable for blacks; (ii) Upward mobility (P (Child in Q5| Parent in Q1)): the* ***lower*** *the racial gap, more favorable for blacks; (iii) Downward mobility (P (Child in Q1| Parent in Q5)): the* ***higher*** *the racial gap, more favorable for blacks; (iv) Remaining rich (P (Child in Q5| Parent in Q5)): the* ***lower*** *the racial gap, more favorable for blacks. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.*

Relatively good students

The last educational factor studied is the student's relative performance to their peers. That is, we are not precisely analyzing the effect of students' grades on intergenerational mobility, but rather the effect on the mobility of students getting good grades compared to the rest of their class. These are two different analyses because, for the first examination, we should rank the entire population of 12th-grade students according to their SARESP scores, but we instead rank them by the school.

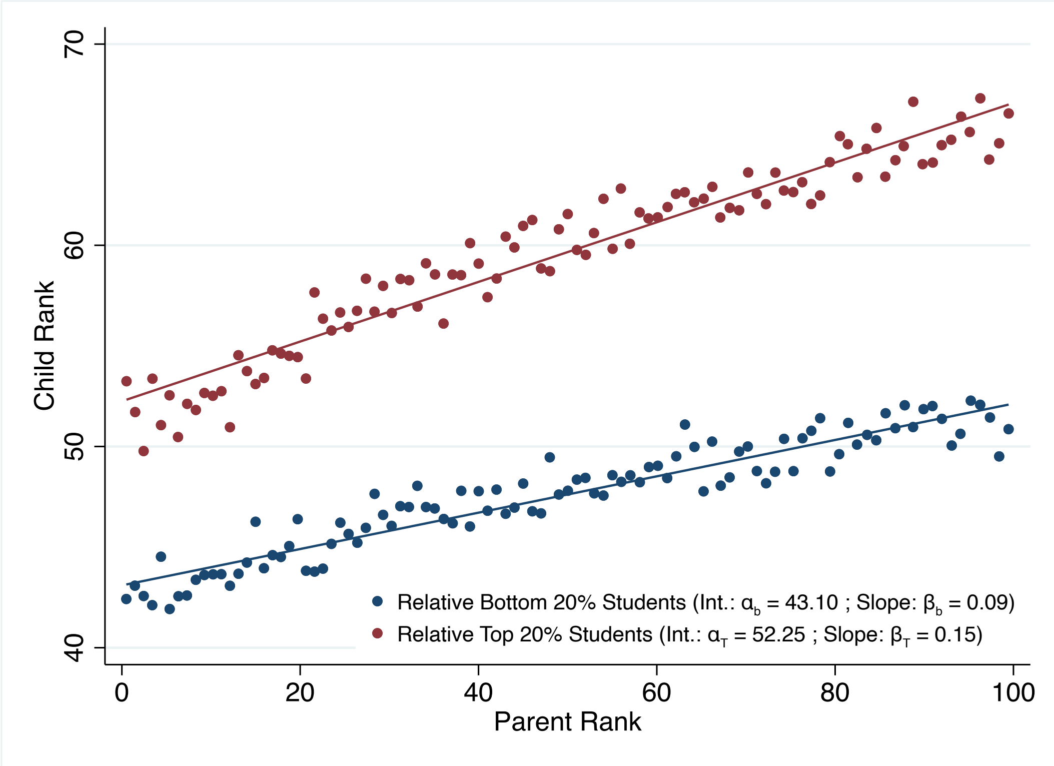
More specifically, for this analysis, we first take all students in the 12th grade from all years we have SARESP data, and we rank the top and bottom students per school based on their SARESP standardized test scores. Some of these students will be in final Sample 1.[12] Therefore, the results will be of the top and bottom students from each school found in each final population (see Section 4, "Variable definition", for more details on this variable's construction). We will name these students relatively top and relatively bottom since they have good or bad test scores compared only to the students in their own school and not the entire population.[13]

Figure 15 plots the children's mean salary rank against their parents' mean income rank for these two groups of students. Note that the bottom students have uniformly lower rates of absolute mobility across the entire parental income distribution. For instance, relatively low-grade students with parents in the 25th percentile reach a salary rank of 46.9 on average, almost 10 percentiles on average below relatively high-grade students born to parents in the same percentile. This discrepancy persists even at the highest levels of income. The difference at the 100th percentile is, on average, 15.5. These results suggest that education positively affects children's future salaries and that even students from high-income families are negatively affected if they perform poorly in their school.

[12] Only a few because they are the ones found both in SARESP and RAIS databases (more about the data and the match explained in Section 4).

[13] It is worth mentioning that there is no strong correlation between the group that represent the 20% best on SARESP (considering all schools together), and the group that represents the 20% best on SARESP in each school. In other words, the relatively better students are not the same as the absolutely better ones. The correlation between the two groups is 0.56.

Figure 15 – Empirical Estimates of Intergenerational Mobility and Relatively Top and Bottom students for Sample 1



Notes: Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS).

Table 13 tells the same story as Figure 15 when we look at absolute probability levels, that regardless of race, relatively good students have significantly higher upward mobility and remaining in rich rates and lower downward mobility and poverty trap rates than relatively bottom students.

It is worth noting that the group of relatively good students also includes good students at low-performing schools. That is, we have here students who are not considered top compared to the entire population of 12th-grade students but, when compared only to their class, are perceived as good students. There is this vivid literature that examines the rank effect in school, showing that students' relative position affects subsequent academic performance, and some mechanisms that mediate this rank effect are related to students' motivations and aspirations (Ladant & Bargagli-Stoffi (2022); Elsner & Isphording (2018)). The idea of this argument narrative is that class rank can impact later school achievement through the student's confidence in their intellectual abilities, shaping one's conceptions about one's intellectual worth. Furthermore, another point explored by Ladant & Bargagli-Stoffi (2022) is that, overall, the rank leads to a significant increase on good perception and a significant decrease on bad perception, which affects students' motivations and aspirations.

We can connect our results to this literature discussion by speculating that students' mechanisms of self-confidence and aspirations may also be affecting intergenerational mobility. Moreover, within the aspirations literature, there are also studies showing that teenage career aspirations interact to influence social status attainment and earnings in adulthood (Ashby & Schoon (2010); Genicot & Ray (2017)). However, further research should be conducted to understand the causal effects of these mechanisms on intergenerational mobility since this paper only speculates possible correlations.

The second objective of this subsection is to understand whether the effects of being a relatively good student are reversed in narrowing the existing racial gaps. For this analysis, we need to look at columns (3) and (4) of Table 13, which follow the same logic as the previous tables presented. In column 3 we calculated the racial gap, which is the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the analysis of the racial gap indicator in column 3 (see Section 4, "Variable Definition" for all the different interpretations of this indicator).

Analyzing column (4), note that among the relatively top students, there is a decrease in the racial gap compared to the relatively bottom students in three of the four socioeconomic cases presented. The only situation where we see that being a relatively top student has higher racial gap is in downward mobility. Although individuals are less likely to experience downward mobility by being relatively top students, the gap between blacks and whites, in this case, is still more significant. But in general, being a relatively good student can reduce racial disparities to some extent.

Following the recent literature, this result also intrigues us to question the effects of relative student rank on reducing racial differences. To the best of our knowledge, no studies explore the causal effect of student rank from this perspective. Therefore, we invite future research to examine the causality of how student relative position impacts racial gaps.

Table 13 – Intergenerational Mobility by Race for different Relatively Top and Bottom Students for Sample 1

	(1) White	(2) Black	(3) = (1)/(2) Racial Gap	(4) Interpretation of (3)
Panel A. Poverty Trap: P(Q1|Q1)				
Relatively Top Students	14.10%	13.77%	1.02	"Relatively Top Students"
Relatively Bottom Students	18.02%	18.78%	0.96	favor blacks more
Panel B. Upward Mobility: P(Q5|Q1)				
Relatively Top Students	20.76%	18.14%	1.14	"Relatively Top Students"
Relatively Bottom Students	8.01%	6.81%	1.18	favor blacks more
Panel C. Downward Mobility: P(Q1|Q5)				
Relatively Top Students	6.75%	8.44%	0.80	"Relatively Bottom Students"
Relatively Bottom Students	12.17%	13.45%	0.90	favor blacks more
Panel D. Remaining Rich: P(Q5|Q5)				
Relatively Top Students	40.34%	35.71%	1.13	"Relatively Top Students"
Relatively Bottom Students	17.16%	13.07%	1.31	favor blacks more
Number of observations	*White*	*Black*		
Relatively Top Students	*49,966*	*29,483*		
Relatively Bottom Students	*39,991*	*39,463*		

*Notes: This table presents the results of the extremes of the transition matrix for Relatively Top and Bottom Students by race, for Sample 1. Sample 1: Parents from children in the 12th grade (SARESP), and these same children when they were 24 to 27 years old in the formal labor market (RAIS). The relatively top and bottom students variable was calculated by averaging the Portuguese and Mathematics scores that each student took in the SARESP test in 12th grade, then we ranked the scores by school, and created a dummy that assumes one if the student belongs to the top 20% of students in that school in that grade, and zero if the student belongs to the bottom 20% of students in that school in that grade. This variable is a proxy for measuring whether a student is a relatively good-grade student or a relatively bad-grade student in their school context. The racial gap is calculated as the ratio between the white's probability (column 1) and the black's probability (column 2). The interpretation of (3) refers to the interpretation of the racial gap indicator in column 3. This indicator measures how much whites' probability of being in the respective socioeconomic situation (poverty trap, upward or downward mobility, or remaining rich) is higher (> 1) or lower (< 1) than blacks'. This indicator should be interpreted with caution because the interpretation changes in each socioeconomic situation, so we have already added column (4) with the conclusion of its interpretation. We consider better the educational setting that is more favorable for the blacks and therefore is going in the direction of reducing racial differences. The interpretation for each row should be made as follows: (i) Poverty trap situation (P (Child in Q1| Parent in Q1)): the **higher** the racial gap, more favorable for blacks; (ii) Upward mobility (P (Child in Q5| Parent in Q1)): the **lower** the racial gap, more favorable for blacks; (iii) Downward mobility (P (Child in Q1| Parent in Q5)): the **higher** the racial gap, more favorable for blacks; (iv) Remaining rich (P (Child in Q5| Parent in Q5)): the **lower** the racial gap, more favorable for blacks. Q1 and Q5 refer to the first and fifth quintiles of the income distribution, respectively.*

6 Conclusion

This research paper documents how school and educational factors relate to intergenerational mobility for different racial groups. We use unique data from the state of São Paulo that connects longitudinally two generations, precisely, information on the parent's socioeconomic conditions to the children's income in the formal labor market. This study does not propose causal relationships, however, it brings compelling discussions by relating literature on education and racial inequality to assess intergenerational mobility.

We find that even in one of the most wealthy states in Brazil, there is a strong intergenerational persistence of income: an individual born into a more socially advantaged family is, on average, two times more likely to be among the top 25% of earners as a young adult. Furthermore, we find a significant difference between the intergenerational mobility of whites and blacks.

With regard to school and educational factors, we find that what happens in students' trajectory between 9th and 12th grade has a significant impact on relative immobility. During this period, the parent's socioeconomic status becomes very determinant in their children's future income. Thus, this seems to be a critical time to guarantee compensatory policies that give the low-income students future income trajectories more similar to the high-income ones. Therefore, future research should be conducted to understand the critical aspects of the low and high-income individuals' educational trajectories influencing this relative immobility during high school.

We follow by analyzing the 9th-grade student's educational trajectory and high and low-performance schools. High school graduates have higher absolute and relative mobilities than those who drop out. For instance, children with a high school diploma are, on average, 43% more likely to experience upward mobility than students who dropped out of school. In addition, schools with higher high-stakes test scores increase the probability of upward mobility for all students, but the wealthier seem to appropriate the benefits of school performance more. Therefore, attending schools with good performance in high-stake exams alone does not seem sufficient to lift individuals out of the poverty trap. Furthermore, although having a high school diploma and higher-scoring schools favor black individuals more than the other educational contexts (dropout school or low-performance schools) for upward mobility and remaining rich, they do not seem to be able to prevent blacks from falling socially or keeping them in poverty. This latter evidence also fosters ideas for future research agendas, which could ascertain why blacks are only favored in these two socioeconomic settings, for example.

In terms of social and racial segregation, we find that students from all social strata benefit when they are among wealthier peers. On the contrary, schools with a high concentration of wealthy parents are associated with higher racial gaps. Moreover, schools with higher racial diversity are associated with higher future earnings. However, racial integration in schools alone seems insufficient to close the existing racial gap.

Finally, we discover that students who get good test scores in their school environment are associated with higher future absolute gains than their peers. Moreover, these gains seem to favor blacks more relatively than whites, narrowing the racial gap for future generations. Here there is room to causally relate this recent literature about the effects on students' motivation and aspiration according to their relative rank in schools with the intergenerational mobility literature.

Throughout this paper, we conjecture possible causes and consequences for the correlations found between educational factors and intergenerational mobility for different racial groups. Although these associations are based on the recent literature on education, racial inequality, and intergenerational mobility, they do not have a causal nature. Thus, for these conjectures to effectively help implement more assertive public policies to reduce intergenerational racial inequalities or to increase student mobility, a more in-depth causal research agenda is needed on the suggested topics.

References

ARIAS, O.; YAMADA, G.; TEJERINA, L. Education, family background and racial earnings inequality in brazil. International journal of manpower, Emerald Group Publishing Limited, v. 25, n. 3/4, p. 355–374, 2004.

ASHBY, J. S.; SCHOON, I. Career success: The role of teenage career aspirations, ambition value and gender in predicting adult social status and earnings. Journal of vocational behavior, Elsevier, v. 77, n. 3, p. 350–360, 2010.

BERTRAND, M.; MULLAINATHAN, S. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. American economic review, v. 94, n. 4, p. 991–1013, 2004.

BOTELHO, F.; MADEIRA, R. A.; RANGEL, M. A. Racial discrimination in grading: Evidence from brazil. American Economic Journal: Applied Economics, v. 7, n. 4, p. 37–52, 2015.

BRITTO, D. G. et al. Intergenerational mobility in the land of inequality. BAFFI CAREFIN Centre Research Paper, n. 2022-186, 2022.

BRUNELLO, G.; CHECCHI, D. School quality and family background in italy. Economics of education review, Elsevier, v. 24, n. 5, p. 563–577, 2005.

CHANCEL, L.; PIKETTY, T.; SAEZ, E. World inequality report 2022. [S.l.]: Harvard University Press, 2022.

CHETTY, R. et al. Race and economic opportunity in the united states: An intergenerational perspective. The Quarterly Journal of Economics, Oxford University Press, v. 135, n. 2, p. 711–783, 2020.

CHETTY, R. et al. Is the united states still a land of opportunity? recent trends in intergenerational mobility. American Economic Review, v. 104, n. 5, p. 141–47, 2014.

CHETTY, R. et al. Where is the land of opportunity? the geography of intergenerational mobility in the united states. The Quarterly Journal of Economics, Oxford University Press, v. 129, n. 4, p. 1553–1623, 2014.

CHETTY, R. et al. Social capital i: measurement and associations with economic mobility. Nature, Nature Publishing Group, v. 608, n. 7921, p. 108–121, 2022.

CORAK, M.; HEISZ, A. The intergenerational earnings and income mobility of canadian men: Evidence from longitudinal income tax data. Journal of Human Resources, JSTOR, p. 504–533, 1999.

DAHL, M. W.; DELEIRE, T. The association between children's earnings and fathers' lifetime earnings: estimates using administrative data. [S.l.]: University of Wisconsin-Madison, Institute for Research on Poverty Madison ..., 2008.

DOUGLAS, S.; MASSEY, D. American apartheid: Segregation and the making of the underclass. [S.l.]: Harvard University Press Cambridge, MA, 1993.

DURLAUF, S. N. et al. Poverty traps. [S.l.]: Princeton University Press, 2006.

ELSNER, B.; ISPHORDING, I. E. Rank, sex, drugs, and crime. Journal of Human Resources, University of Wisconsin Press, v. 53, n. 2, p. 356–381, 2018.

FERGUSON, B. D. et al. Estimating permanent income using indicator variables. Health systems performance assessment: debates, methods and empiricism. Geneva: World Health Organization, Citeseer, p. 747–60, 2003.

FILMER, D.; PRITCHETT, L. H. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of india. Demography, Duke University Press, v. 38, n. 1, p. 115–132, 2001.

FIRPO, S.; FRANÇA, M.; PORTELLA, A. Racial inequality in the brazilian labor market and the role of education. Available at SSRN 3967828, 2021.

FLUG, K.; SPILIMBERGO, A.; WACHTENHEIM, E. Investment in education: do economic volatility and credit constraints matter? Journal of Development Economics, Elsevier, v. 55, n. 2, p. 465–481, 1998.

GENICOT, G.; RAY, D. Aspirations and inequality. Econometrica, Wiley Online Library, v. 85, n. 2, p. 489–519, 2017.

GRAWE, N. D. Lifecycle bias in estimates of intergenerational earnings persistence. Labour economics, Elsevier, v. 13, n. 5, p. 551–570, 2006.

IBGE. Rendimento de todas as fontes 2021 (PNAD) - Informativo. [S.l.], 2022.

KRAAY, A.; MCKENZIE, D. Do poverty traps exist? assessing the evidence. Journal of Economic Perspectives, v. 28, n. 3, p. 127–48, 2014.

LADANT, P. S. F.-X.; BARGAGLI-STOFFI, F. "rather first in a village than second in rome? the effect of students' class rank in primary school on subsequent academic achievements [job market paper]". Working Paper., 2022.

LIN, N.; DUMIN, M. Access to occupations through social ties. Social networks, Elsevier, v. 8, n. 4, p. 365–385, 1986.

LOCHNER, L.; MONGE-NARANJO, A. Credit constraints in education. Annu. Rev. Econ., Annual Reviews, v. 4, n. 1, p. 225–256, 2012.

LOURY, G. A Dynamic Theory of Racial Income Differences, w: Women, Minorities, and Employment Discrimination,(eds.) Wallace PA, Le Mond AM. [S.l.]: Lexington Books, Langham, 1977.

MADEIRA, R. A.; RANGEL, M. A. Racial achievement gaps in another america: Discussing schooling outcomes and affirmative action in brazil. In: closing the Achievement Gap from an international perspective. [S.l.]: Springer, 2014. p. 127–160.

MCKENZIE, D. J. Measuring inequality with asset indicators. Journal of population economics, Springer, v. 18, n. 2, p. 229–260, 2005.

MURPHY, K. M.; WELCH, F. Empirical age-earnings profiles. Journal of Labor economics, University of Chicago Press, v. 8, n. 2, p. 202–229, 1990.

QUILLIAN, L. et al. Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 114, n. 41, p. 10870–10875, 2017.

SAHN, D. E.; STIFEL, D. Exploring alternative measures of welfare in the absence of expenditure data. Review of income and wealth, Wiley Online Library, v. 49, n. 4, p. 463–489, 2003.

SOLON, G. Intergenerational mobility in the labor market. In: Handbook of labor economics. [S.l.]: Elsevier, 1999. v. 3, p. 1761–1800.

TELLES, E. E. Race in another america. In: Race in Another America. [S.l.]: Princeton University Press, 2014.

TORCHE, F. Intergenerational mobility and gender in mexico. Social Forces, University of North Carolina Chapel Hill, v. 94, n. 2, p. 563–587, 2015.

VYAS, S.; KUMARANAYAKE, L. Constructing socio-economic status indices: how to use principal components analysis. Health policy and planning, Oxford University Press, v. 21, n. 6, p. 459–468, 2006.

Appendix A. Correlation between socioeconomic indicators

Table A.1 – Correlation between socioeconomic indicators estimated by different methods

Panel A. 9th Grade

	PCA and Correlation Matrix	**PCA and Covariance Matrix**	**Factor Analysis**
PCA and Correlation Matrix	1.0	-	-
PCA and Covariance Matrix	0.9799	1.0	-
Factor Analysis	0.9793	0.9886	1.0

Panel B. 12th grade

	PCA and Correlation Matrix	**PCA and Covariance Matrix**	**Factor Analysis**
PCA and Correlation Matrix	1.0	-	-
PCA and Covariance Matrix	0.9785	1.0	-
Factor Analysis	0.9777	0.9806	1.0

Notes: Correlation between predicted values calculated using three different methods: (i) PCA with correlation matrix; (ii) PCA with covariance matrix and; (iii) Factor Analysis (Principal Factor). The predicted values correspond to the socioeconomic indicator.

Appendix B. Rank-rank regression by race

Table B.2 – Rank-rank regression by race

	Children Percentile
Parent percentile	0.1141***
	(0.0021)
White × Parent Percentile	0.0204***
	(0.0028)
White	1.2857***
	(0.1602)
Constant	45.7502***
	(0.1138)
Observations	388,682
	P-Value
F-test constant	0.000
F-test slope	0.000

*Notes: The table reports the rank-rank regression's estimated constant (absolute mobility) and slope (relative mobility) by race. The Parent's Percentile is the parent's ranking. We rank parents based on their Wealth Index relative to all other parents in 12th grade. The Children's Percentile corresponds to the children's ranking. We rank children based on their income relative to all other workers aged 24 to 27 from RAIS. Our primary measure for children's income is the December median deflated earnings when children are between 24 and 27. The deflation was calculated using the consumer price index (IPCA, in Portuguese) with the base year of 2018. White is a dummy that assumes one if the individual is white and zero if the individual is black. The standard deviation is in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The F-test constant corresponds to the F-test for testing whether the absolute mobility between whites and blacks is equal. The F-test slope corresponds to the F-test for testing whether the relative mobility between whites and blacks is equal.*

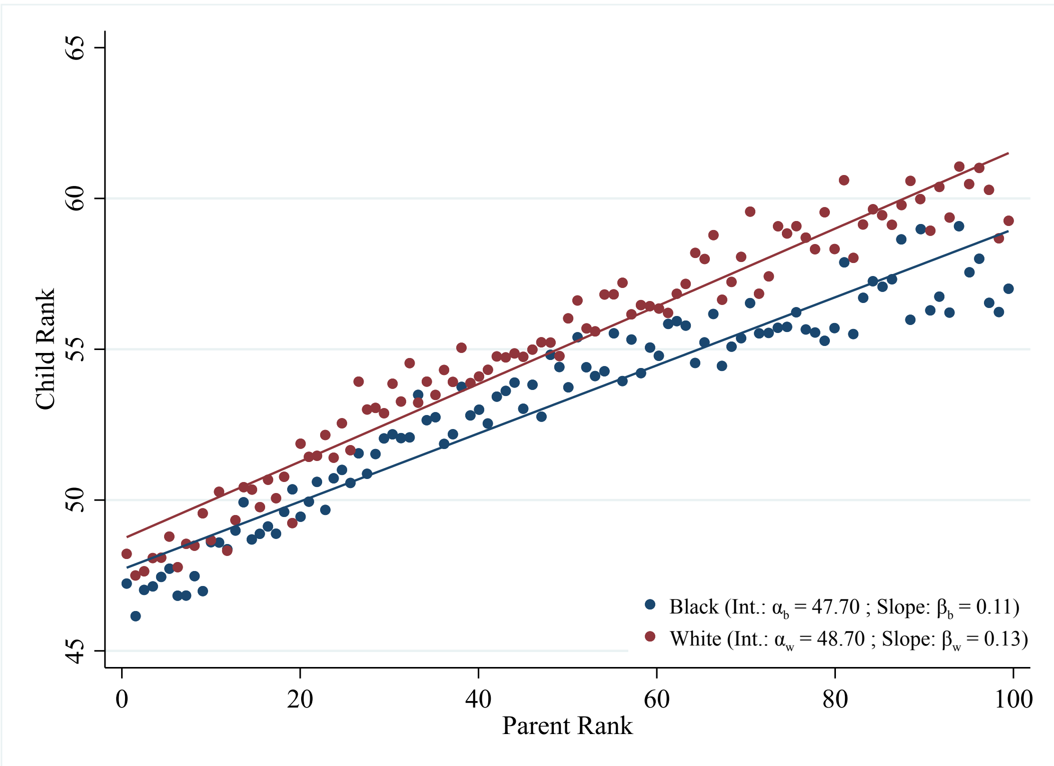
Appendix C. Results of Samples 2 and 3

This section shows the estimated results for Samples 2 and 3. Note that the results presented here are very similar to the ones from Sample 1, and the discussions and main findings are the same. The results that varied from those of Sample 1 exposed are also discussed in the main text.

Blacks and Whites

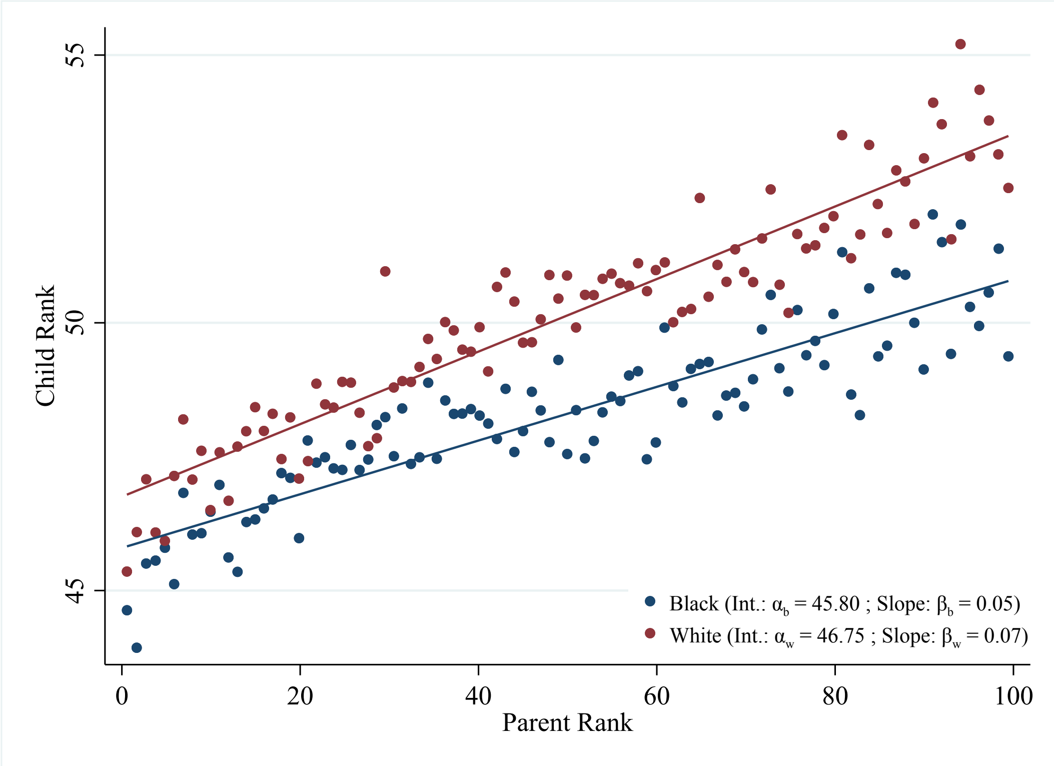
The results below are analogous to Figure 8 for the other working samples (2 and 3).

Figure C.1 – Empirical Estimates of Intergenerational Mobility and Racial Disparities for Sample 2



Notes: Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Figure C.2 – Empirical Estimates of Intergenerational Mobility and Racial Disparities for Sample 3

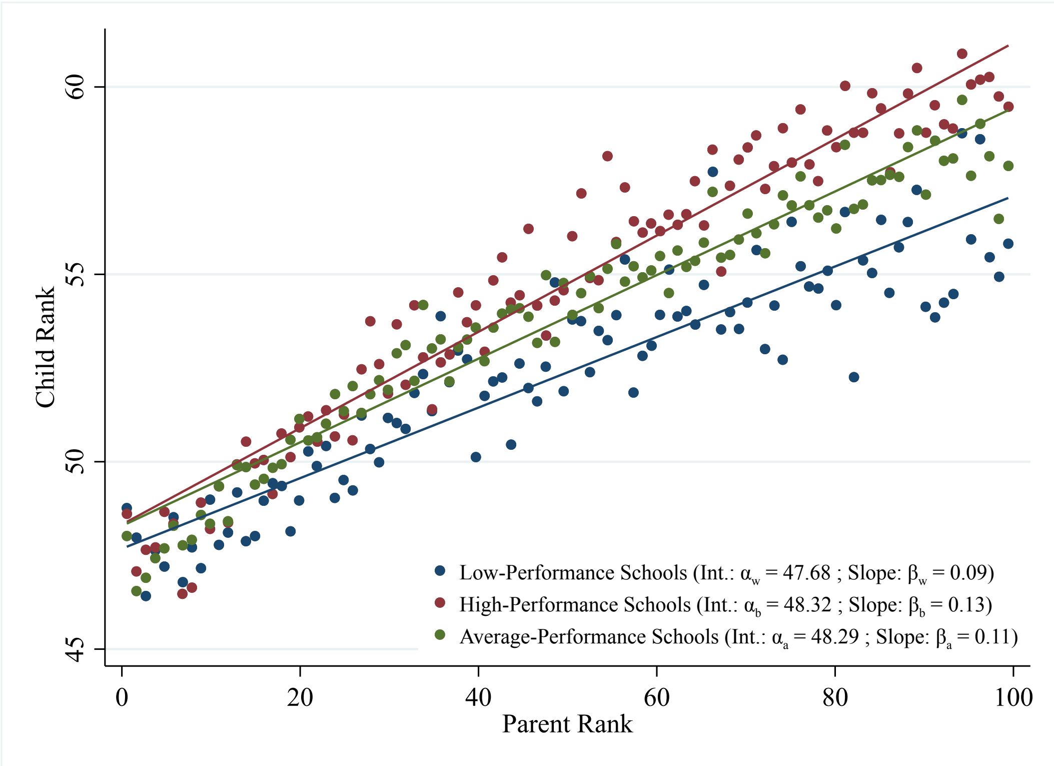


Notes: Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

School Performance

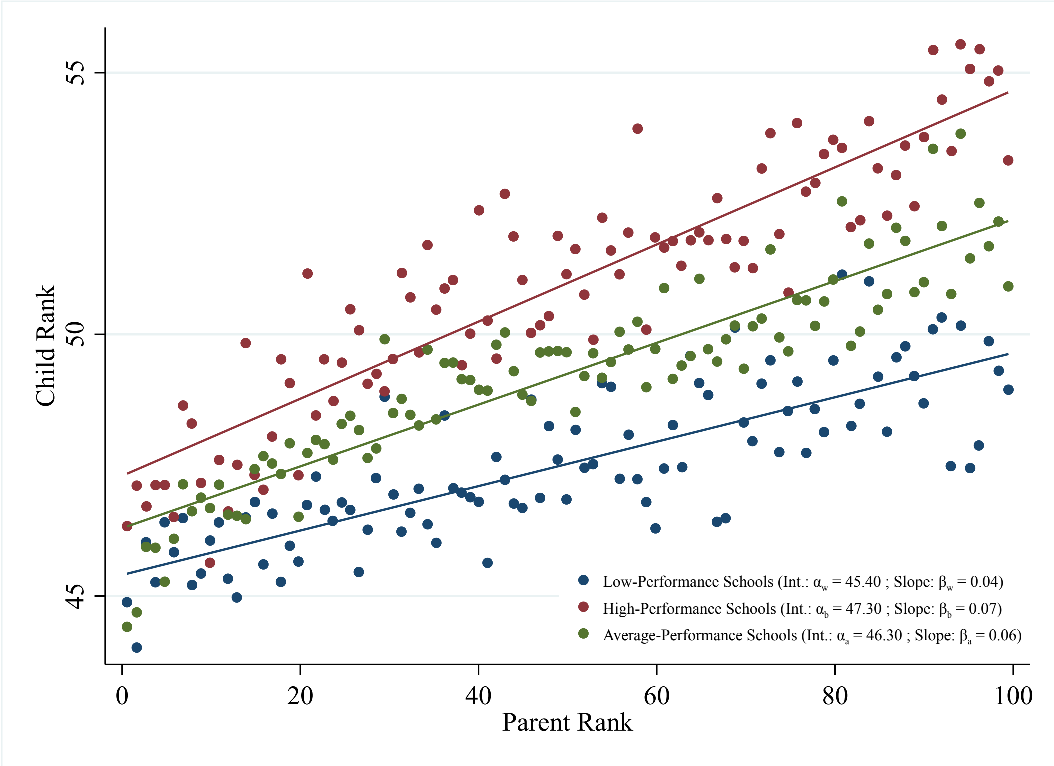
The results below are analogous to Figure 10 for the other working samples (2 and 3).

Figure C.3 – Empirical Estimates of Intergenerational Mobility for Different School Performance Categories for Sample 2



Notes: Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Figure C.4 – Empirical Estimates of Intergenerational Mobility for Different School Performance Categories for Sample 3

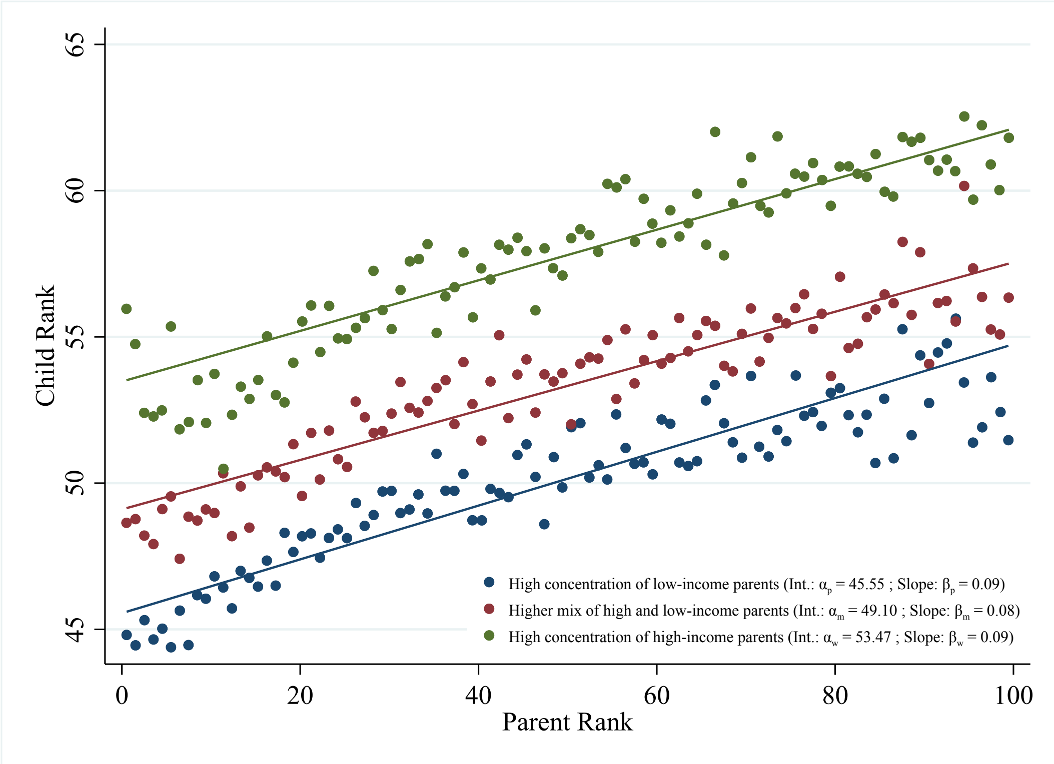


Notes: Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Social Segregation in School

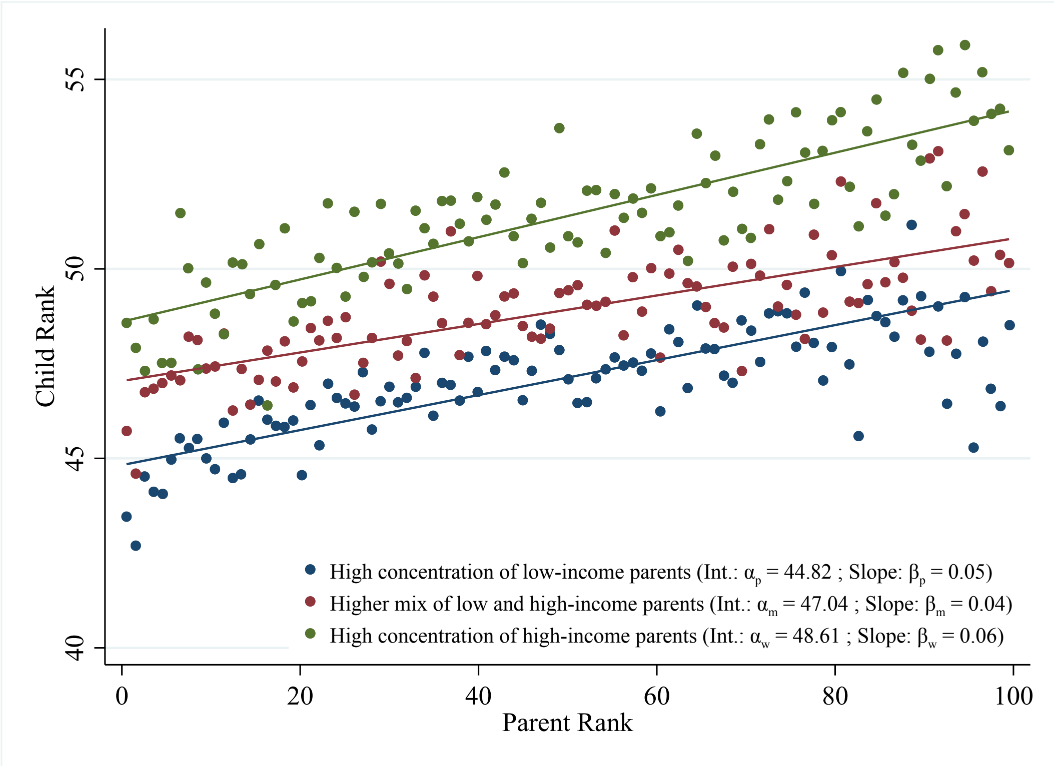
The results below are analogous to Figure 12 for the other working samples (2 and 3).

Figure C.5 – Empirical Estimates of Intergenerational Mobility for Different School Categories, in terms of social inequality for Sample 2



Notes: Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Figure C.6 – Empirical Estimates of Intergenerational Mobility for Different School Categories, in terms of social inequality for Sample 3

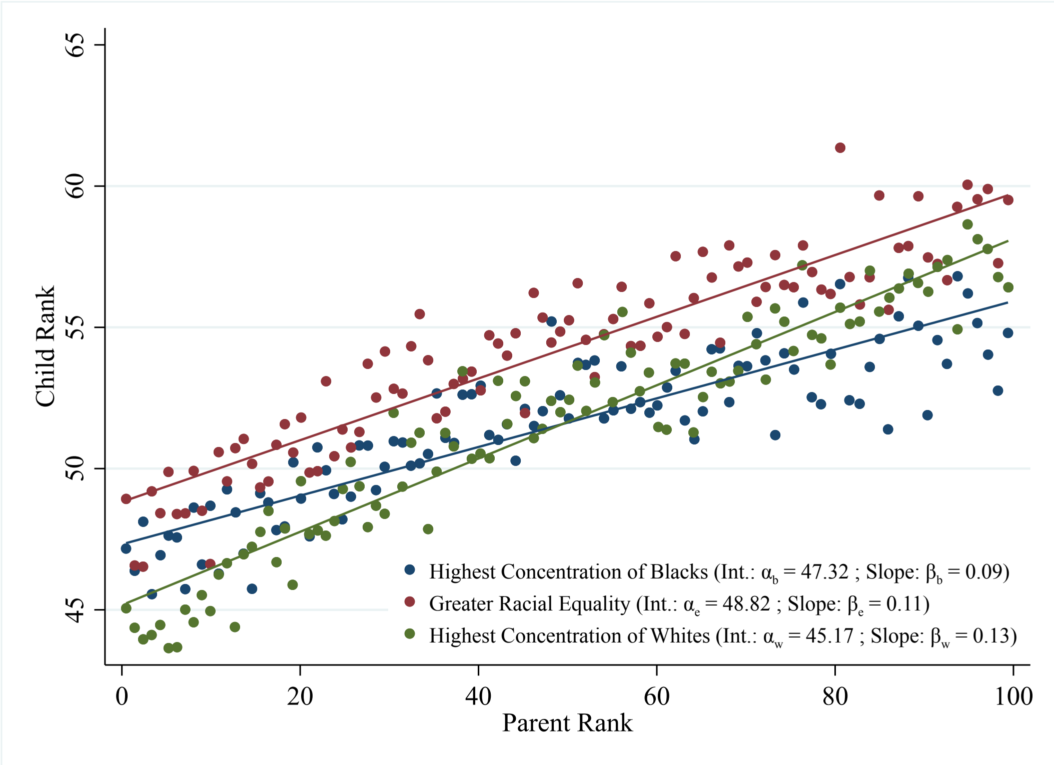


Notes: Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Racial Segregation in School

The results below are analogous to Figure 13 for working Sample 2.

Figure C.7 – Empirical Estimates of Intergenerational Mobility for Different School Categories, in terms of racial concentration for Sample 2

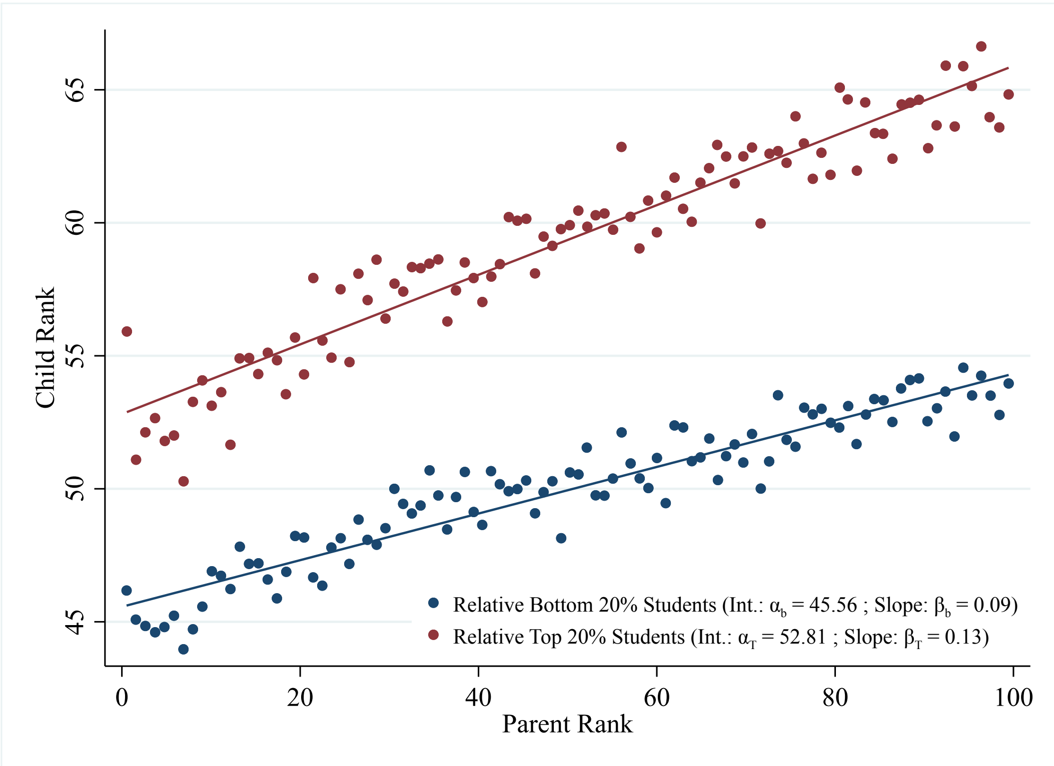


Notes: Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Relatively good students

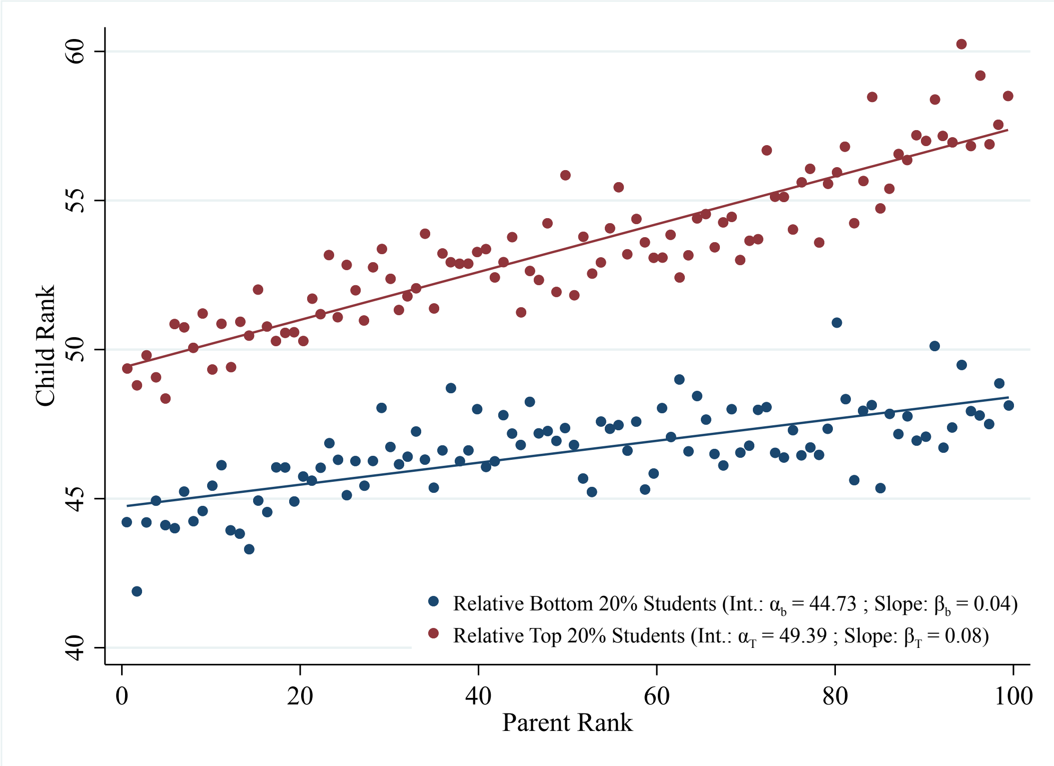
The results below are analogous to Figure 15 for the other working samples (2 and 3).

Figure C.8 – Empirical Estimates of Intergenerational Mobility and Relatively Top and Bottom students for Sample 2



Notes: Sample 2: Parents from children in the 12th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).

Figure C.9 – Empirical Estimates of Intergenerational Mobility and Relatively Top and Bottom students for Sample 3



Notes: Sample 3: Parents from children in the 9th grade (SARESP), and these same children when they were 20 to 24 years old in the formal labor market (RAIS).