Impacts of grade configuration on Brazilian student outcomes[*]

Marcos Vinicio Wink Junior[†]
Luis Henrique Zanandréa Paese[‡]
Marcelo de Carvalho Griebeler[§]

Contents

Abstract · Resumo

In this paper we evaluate the impact of grade span configuration on student outcomes. We build a model which shows that the more homogeneous the cohort of students the higher the share of them which achieves the minimum level of academic performance required in an exam. We then test this theoretical finding by comparing performances of 5th-grade students of elementary schools (1st to 5th grade) and those of elementary-middle schools (1st to 9th grade) in Brazil. PSM and PSM with Diff-in-Diff methodologies are used to control possible biases. We find that elementary schools present better results in Portuguese language and mathematics standardized score tests, higher passing rates and lower dropout rates than elementary-middle schools, which corroborates the theoretical results. The robustness of estimates is checked through several tests and alternative specifications. Finally, we found evidence that the main mechanisms behind our results are related to alternative pedagogical practices and school management policies.

1. Introduction

Over the last decades, Brazil has been facing a profound change in its demographic structure. While the number of elderly people is increasing, the number of children has been falling due to the fertility rate reduction. Although this change in the

[†]Universidade do Estado de Santa Catarina (UDESC), Centro de Ciências da Administração e Socioeconômicas (Esag), Departamento de Ciências Econômicas. Avenida Madre Benvenuta 2037, Itacorubí, Florianópolis, SC, CEP 88035-001, Brasil. 0000-0003-2206-5211

[‡]Centro Administrativo Sicredi. Avenida Assis Brasil, Jardim Lindoia, Porto Alegre, RS, CEP 91060-900, Brasil. 0000-0002-8639-9459

[§]Universidade Federal do Rio Grande do Sul (UFRGS), Faculdade de Ciências Econômicas, Departamento Economia e Relações Internacionais. Avenida João Pessoa 52, Centro Histórico, Porto Alegre, RS, CEP 90040-000, Brasil. 0000-0001-7943-4802

✉ marcos.winkjunior@udesc.br   ✉ luishzp@hotmail.com   ✉ marcelo.griebeler@ufrgs.br

age pyramid is a macroeconomic challenge in the short term, it can provide, for example, opportunities to increase the government expenditure on education per student. Such a demographic transition may also create incentives to modify the number of schools supplied to the public. Figure 1 shows the enrollment reduction in elementary and middle school in the last 10 years. It is important to note that the attendance rate did not decrease during this period.

Considering the Brazilian demographic phenomenon, the question regarding what to do with schools that have few students enrolled arises. The first and simplest answer to this question would be to close these schools and relocate their students. The second possibility would be to change the school grade span configuration by separating the elementary school (1st to 5th grade) from the middle school (6th to 9th grade). The latter proposal has been considered in Brazil during the last years.[1] This issue became even more debated in Brazil with the publication of Decree No. 61672, of November 30, 2015, by the Government of the State of São Paulo. The decree authorized the process of grade span configuration for state public schools. This proposal, however, was revoked in May 2016 (Decree No. 61962) due both to the absence of studies measuring the possible impacts of this policy on Brazilian students and public pressure.

The main supporting argument of this policy is that the school reorganization would be beneficial since it would be providing more specific pedagogical plans, by increasing the specialization of principals and teachers. In other words, principals
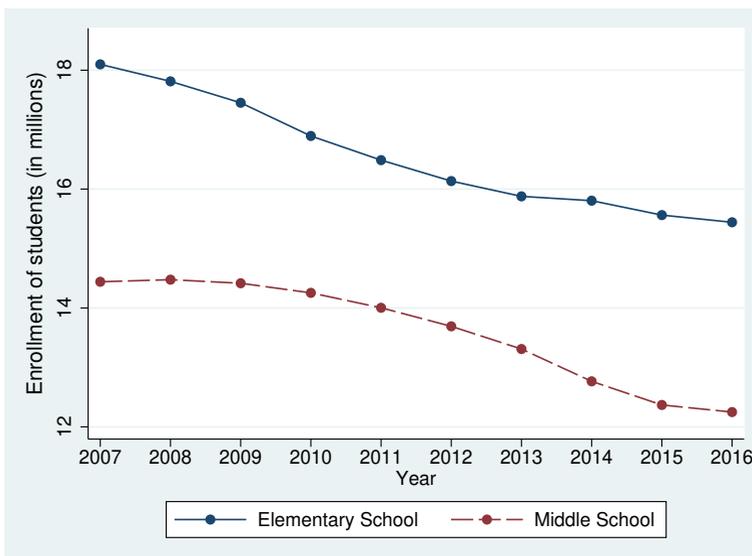

Figure 1. Total Enrollment of Students in Elementary and Middle School in Brazil, 2007–2016.

[1]In order to have an idea of the importance of each school type, observe that according to the Scholar Census of 2015, elementary schools represent 40% of the total number of schools in Brazil.

and teachers may become more likely to use alternative practices, focused on the specific needs of students. In this context, principals also may be more successful in conducting management that benefits teachers. The channel by which such a policy would affect teachers and principals behaviors is the higher level of homogeneity of the cohort, given that it is expected that these schools will have students with more homogeneous characteristics. In fact, empirical evidence has shown that heterogeneous classes negatively affect students' individual achievements (Hanushek, Kain, Markman, & Rivkin, 2003; Hattie, 2002). Grade span configuration may also play an important role in class size reduction. Although there is still an empirical discussion on this subject, some recent studies have found that increases in class size have negative impact on students' achievement. Moreover, the effect of class size reducing is larger for low-income children (Schanzenbach, 2014; Bosworth, 2014).

A process similar to the one of the Government of São Paulo occurred in the United States in the mid-20th century. Nowadays, most American students switch from elementary to middle school before entering high school. The effects of this transition have been widely studied empirically and evidence that switching schools affects school outcomes negatively has been found. Looking thoroughly at academic achievement, several studies have shown that this transition negatively affects mathematics and reading scores (Byrnes & Ruby, 2007; Offenberg, 2001; Rockoff & Lockwood, 2010; Dhuey, 2013; Holmlund & Böhlmark, 2019). The literature also shows that switching school affects non-academic outcomes (Schwerdt & West, 2013; Weiss & Bearman, 2007). The most discussed mechanism behind these findings is the transition period when students are teenagers and several behavioral changes are taking place simultaneously.

The long-term effects of changing school, however, are not considered by the aforementioned literature. The potential benefits of the grade span configuration on academic achievement may not be immediate. This is documented, for example, by Dove, Pearson, and Hooper (2010), which investigated the effects of this policy on school achievement measured by the Arkansas Benchmark Examination for sixth-grade students. The study reveals that there are positive effects on mathematics proficiency tests after the second year of transition. Yet, there is no evidence of effects on the literacy achievement.

There are two factors common to aforementioned studies, however. The first is that they only analyze the case of developed countries. In fact, the literature does not provide any similar work which studies grade span configuration and its impacts on students' performance in developing countries. The second is that they evaluate students only after the transition to middle school. Once again, to the best of our knowledge, there are no studies that investigate the possible positive effects for early students of elementary school specialization before transition. Those characteristics suggest that there are important literature gaps, which we expect to explore in this paper.

Brazil is an interesting case because teachers are not specialized to teach in specific series. According to data from SAEB 2015, 40% of 5th grade teachers have more than one job, 65% have been teaching in this grade for less than 5 years and 57% have no degree in education or pedagogy. Since there is usually only one teacher per class in 5th grade, the school reorganization could compensate the lack of specific training with learning by doing. In fact, with a more homogeneous cohort—in terms of age and grade—, teachers could specialize by teaching the same grade repeatedly.

In this paper we use the largest education dataset in Brazil to measure the effect of grade span configuration on 5th grade students enrolled in the Brazilian Public School System. We focus on four different outcomes: achievement in mathematics and reading, passing and dropout rates. We find that both achievements and passing rates are significantly higher for elementary schools, when compared to elementary-middle ones. We also found that the dropout rate is smaller for the elementary schools. In order to estimate the causal effects, we employed the Propensity Score Matching Strategy to compare outcomes from 5th grade students enrolled in elementary and elementary-middle schools. In other words,we compare the performance of 5th grade students in schools that only provide the first 5 years of primary school with the performance of 5th grade students attending schools that provide the full 9 years of primary school. Thus, the contribution of this paper to the literature is twofold: it evaluates the effects of grade span configuration on student's outcomes before their transition in a country different from the USA (a developing one) and also investigates the theoretical and empirical mechanisms that explain those effects.

The empirical exercise we perform is based on the predictions of a game-theoretic model which also provide the mechanisms for the econometric results we find. By assuming that both teachers and principal seek to implement the pedagogy which is the closest to the student's ideal, our model predicts that the more homogeneous the cohort of students, the higher the share of them which achieve a good performance. Such an effect is explained by the decrease in the principal's marginal cost of effort—as a consequence of the higher homogeneity—, which makes him exert more effort. This makes teachers exert more effort as well, which in turn increases students' performance.

We also investigate the possible channels through which school specialization may improve school outcomes in 5th grade students. Among those tested, we find that alternative pedagogical practices and school management policies have important and significant impact. Those results indicate that principals of Brazilian elementary schools seem to diversify pedagogical practices and conduct a management that benefits teachers. Thus, we are able to confirm the main conclusions of our game-theoretic model. We believe that the findings of this study generate significant impacts for future public policy proposals. This paper also shed some light on the ongoing debate in São Paulo, despite further studies being needed.

The paper proceeds as follows. In section 2 we present the game theoretical model, where we analyze the channels through which the change in the grade configuration may affect student's performance. Section 3 describes the Brazilian data used in the empirical exercise. Section 4 explains the methodology used. Our main findings are described in section 5. In section 6 we present some tests to check the robustness of our results and section 7 concludes. The omitted propositions' proofs can be found in Appendix A.

2. Model

A school is composed by a principal and a continuum of identical teachers of mass 1. Each teacher is in charge of exactly one class, such that there is a continuum of classes of mass 1 as well. Each class is composed by a continuum of students of mass 1, who are heterogeneous both within and between classes. We model this by assuming that the learning process varies among the students, such that there is an "ideal pedagogy" for each one of them. Let $\theta_{ij} \in \mathbb{R}$ be such a pedagogy for student i in the class j, then assume that for all class j the distribution of ideal pedagogies of students satisfies $\theta_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. We denote $F_j(\cdot)$ and $f_j(\cdot)$ as the cumulative distribution and the density functions of the normal distribution of class j, respectively. Thus, each class can be characterized by the pair $(\mu_j, \sigma_j^2) \in \mathbb{R} \times [0, \infty)$.

The heterogeneity of classes in the school is summarized by the random variable $\mu \in \mathbb{R}$, which also follows a normal distribution, such that we formally have $\mu \sim \mathcal{N}(\mu^*, \sigma_*^2)$.[2] We denote $\Phi(\cdot)$ and $\phi(\cdot)$ as the cumulative distribution and the density functions of μ. Our main result investigates the impact of changes in the variability of μ on the performance of students of the school. The idea is that schools whose students are more homogeneous are able to become more specialized and thus improve their pedagogical practices, which in turn will affect students' performance positively. The only measure of the student performance we consider is an exam, in which it is required to achieve at least $w < 0$ in order to be able to go to the next grade. Let p_j be the pedagogy actually implemented by the teacher in the class j,[3] then the score of student i is $s_{ij} = -(\theta_{ij} - p_j)^2$.

The pedagogy implemented in each class j depends directly only on the effort y_j of the teacher. The pedagogy function $p : [0, \infty) \to \mathbb{R}$ has the following properties: $p'(\hat{y}_j) = 0$ and $p''(y_j) < 0$ for all $y_j \in [0, \infty)$, such that $p'(y_j) > 0$ if $y_j < \hat{y}_j$; $p'(y_j) < 0$ if $y_j > \hat{y}_j$; and $p(\hat{y}_j) = \mu_j$, where $\hat{y}_j \neq 0$ for all j. The functional form of the pedagogy function is the same for all teachers j, such that their images may be different only because of the different levels of effort teachers choose. The above

---

[2]Observe that the normal distribution of μ is not an assumption. Rather, it follows from the central limit theorem, given that each μ_j has normal distribution as well.

[3]Formally, $p_j := p(y_j)$.

assumptions imply that there exists a level of teacher's effort that provides the ideal pedagogy of the median student, and any deviation of this level causes $p(y_j) < \mu_j$. We can justify that by noticing that, given the heterogeneity of ideal pedagogies and its distribution, if the teacher wants their pedagogy p_j to be close to as many students' ideal pedagogies as possible, his optimal level of effort must be $\hat{y}_j$. A further implication of the above assumptions is that too much effort can be harmful for the class as whole.

2.1 Players

2.1.1 Teachers

All teachers are identical, such that it suffices to analyze the behavior of any member j. As mentioned above, it is reasonable to assume that the teacher wants to provide the pedagogy that is close to the majority of students' ideals. However, in order to achieve it, some effort is required, which yields disutility. We therefore model the utility of the teacher j as

$$u(y_j) = -\mathbb{E}\left[\left(\theta_j - p(y_j)\right)^2\right] - C(y_j, x), \tag{1}$$

where $C(\cdot)$ is a cost function which depends on his own effort and principal's effort x. Once again, we assume that the functional form of the cost function is the same for all teachers, which implies that the utility itself is the same for all j.

The cost function of effort has the following properties: $C_y(\cdot) > 0$; $C_x(\cdot) < 0$; $C_{yy}(\cdot) > 0$; $C_{xx}(\cdot) > 0$; $C_{yx}(\cdot) < 0$; $C(0, x) = 0$ for all $x \in [0, \infty)$; and $C(y_j, 0) > 0$ for all $y_j \in [0, \infty)$. Observe that we assume that the cost is increasing and convex in the teacher's effort, which are standard assumptions. Other standard requirement we make is that the cost associated with zero effort is null. Yet the key characteristics of $C(\cdot)$ is the influence of the principal on it: the higher his effort the lower both the total cost and the marginal cost of the teacher's effort. We also assume that $\lim_{x \to \infty} C_y(y_j, x) = 0$ for all $y_j \in [0, \infty)$, which means that the marginal cost vanishes as the effort of principal increases indefinitely. The intuition is that some policies implemented by the principal—which in turn requires his efforts—in the school may make the teacher's job easier (e.g. teacher training programs and periodically school board meetings). We assume that such an effect is strong, given that it affects both the total and the marginal cost. Finally, we require that the marginal cost of the effort of teacher is null when he chooses zero effort, formally $C_y(0, x) = 0$.

2.1.2 The principal

The preferences of the principal are similar to the teacher's. He wants the pedagogy chosen by the teacher in each class j of his school to be close to as many students' ideal ones as possible. His influence on this variable is indirect: he can implement policies which make less costly the teachers' job. As the implementation of such

policies require effort, he also faces a standard trade-off. Recalling that each class may be characterized by the mean of the ideals pedagogies of its students, the following utility function may represent the principal's preferences:

$$v(x) = -\mathbb{E}\left[\left(\mu - p(y_j)\right)^2\right] - B(x; \sigma_*), \tag{2}$$

where $B(\cdot)$ is the cost function associated with his effort x and σ_* is the standard deviation of the distribution of the mean of ideal pedagogies μ.

   The cost function $B(\cdot)$ has the following properties: $B_x(x, \sigma_*) > 0$ for all $(x, \sigma_*) \in (0, \infty) \times [0, \infty)$ and $B_x(0, \sigma_*) = 0$ for all $\sigma_* \in [0, \infty)$; $B_{xx}(x, \sigma_*) > 0$, $B_\sigma(x, \sigma_*) > 0$, $B_{\sigma\sigma}(x, \sigma_*) > 0$, $B_{x\sigma}(x, \sigma_*) > 0$ for all $(x, \sigma_*) \in [0, \infty) \times [0, \infty)$. Once again, the derivatives with respect to x are standard: the cost associated to the effort of the principal is increasing and convex. However, the crucial characteristics of this function is the impact of the students heterogeneity—measured by σ_*—on $B(\cdot)$. As we argued in the introduction, when students are more homogeneous, the principal is able to adopt policies which make teachers more specialized, which in turn will favor the implementation of better practices and padagogies in classroom by the teachers.

2.2 Timing of the game

This dynamic game with complete information has the following timing:

   1) The principal chooses his level of effort x.
   2) Each teacher j observes it, then decide his optimal level of effort $y_j(x)$.
   3) Students take an exam. Those who score at least w can go to the next grade.

2.3 Equilibrium

In order to solve the game, we must apply backward induction. Thus, let us start by finding the teacher's best response. Each teacher j maximizes (1) by choosing the level of effort y_j. The first order condition (FOC) is given by

$$2p'(y_j)(\mu_j - p(y_j)) = C_y(y_j, x). \tag{3}$$

Observe that $p_j = \mu_j$ only when the marginal cost of effort is null, such that the ideal pedagogy of the median student of the class j never is implemented. In fact, given the properties of $C(\cdot)$, it would be necessary that $y_j = 0$, but in this case $\mu_j > p(y_j)$. Notice that while the left-hand size of (3) measures the marginal benefit of exerting effort, given by the gain of approaching to the median student's ideal position, the right-hand size measures the marginal cost, which is the direct disutility of effort.

   The following proposition analyzes the problem's solution with detail.

Proposition 1. *For each teacher j, there exists a unique optimal level of effort $y_j(x) < \hat{y}_j$ for all $x \in [0, \infty)$. Furthermore, in equilibrium the teacher's optimal effort is an increasing and concave function of the principal's effort, such that we have $y_j'(x) > 0$ and $y_j''(x) < 0$.*

This result is important because it states that the principal can give incentives for the teachers to effort—and thus to improve their students' learning—by himself exerting higher effort. However, this incentive never is strong enough to make teachers achieve the ideal pedagogy of the median student. In fact, the impact of the principal's effort on the teachers' ones increases at a decreasing rate.

Once we have obtained the teachers' best response, we must analyze the principal's maximization problem. Given that he can anticipate that $y_j = y_j(x; \mu_j)$, the FOC of his optimization is

$$2\mathbb{E}\left\{ p'\left(y_j\left(x; \mu_j\right)\right) y_j'\left(x; \mu_j\right) \left[\mu - p\left(y_j\left(x; \mu_j\right)\right)\right] \right\} = B_x(x; \sigma_*). \qquad (4)$$

Notice that, with some abuse of notation, we decide to highlight that μ_j affects the teacher's optimal level of effort y_j. The reason why we choose to do so is that the expectation in (4) is taken over the distribution of μ, namely $\Phi(\mu)$.

Equation (4) represents the trade-off the principal faces when chooses his optimal effort. The left-hand side measures the marginal benefit, which is composed by the impact that his effort has on the teachers' effort and then the indirect effect on approaching the implemented pedagogy to the median student—now, the median student of the whole school. The right-hand side is the marginal cost, which once again is measured by the direct disutility of effort.

Proposition 2. *There exists a unique optimal level of effort $x^* \in (0, \infty)$ for the principal.*

The above result, jointly the one from Proposition 1, allows us to state the equilibrium of the game. In the next section we use it in order to prove the main result of our model.

Corollary 1. *There exists a unique subgame perfect equilibrium of the game above, namely $\left(\left\{ y_j(x^*) \right\}_{j \in \mathbb{R}}, x^* \right)$.*

2.4 More homogeneity better performance?

Our main result states that whenever the school is able to get a more homogeneous cohort, the proportion of students which achieve the minimum score required to go to the next grade is higher. Recall that $s_{ij} = -(\theta_{ij} - p_j)^2$, such that formally that proportion is given by all $\theta_{ij} \in \mathbb{R}$ which satisfy $s_{ij} \geq w$, which in turn can expressed by $S = \int_{p_j - \Delta}^{p_j + \Delta} d\Phi(\mu)$, where $p_j = p(y_j(x^*))$ and $\Delta = \sqrt{-w}$. Now, we are able to state the aforementioned result.

Proposition 3. *The more homogeneous the students in the school the higher the share of them which is able to achieve a score of at least w in the exam.*

The mechanism through which the principal can affect students' performance is clear in the proof of the proposition in Appendix A. Yet the development of the previous section is enough to give us the intuition behind the result. A more homogeneous cohort makes the marginal cost of effort lower for the principal, such that he must then decrease the marginal benefit in order to hold optimality. He does so by decreasing the marginal impact on the teachers' effort, which requires he exerts more effort. As the teachers' best responses are increasing in x, they respond by exerting more effort, which in turn makes the implement pedagogy closer to the one of the median student. Finally, given the assumption that μ is normally distributed, a pedagogy closer to μ^* increases the share of students able to achieve at least w.

3. The dataset

Our main source about schools' types—whether elementary or elementary-middle—is the National Institute for Educational Studies and Research (INEP), an agency subordinate to the Brazilian Ministry of Education (MEC). In order to define which schools are qualified for treatment and control groups, our first task is to gather the number of students enrolled in each grade of every Brazilian school from the 2013 and 2015 waves of the Scholar Census. With regard to information on student's performance and socioeconomic data we use the biennial survey named SAEB (Basic Education Assessment System) and the Scholar Census. SAEB provides information to evaluate the achievement of Brazilian students of 5th and 9th grades. In addition to proficiency tests, SAEB provides information about the social conditions of students, teachers and principals as well as school infrastructure. Even though this survey is applied in a Census form for students enrolled in public schools, a few rules are applied in order to maintain the reliability of the information, i.e. excluding grades with fewer than 20 students enrolled. In order to have a basis of comparison, we only analyze data from public schools in which students were finishing elementary school, that is, 5th graders.[4]

Some remarks about the Brazilian educational system are required to understand how schools are classified between control and treatment groups. First, home-schooling is not legally allowed, such that the Census data we use account every Brazilian student who is enrolled from 1st to 9th grade. Second, according to the UNESCO's International Standard Classification of Education (ISCED), Brazilian Fundamental School (equivalent to regular elementary-middle school) is divided into elementary school, from 1st to 5th grade, and middle school, from 6th to 9th grade.

---

[4]Which represents 98% of the total enrollments of the 5th year.

Therefore, there are three different school types in Brazil: those two defined by ISCED (elementary and middle), and elementary middle school, from 1st to 5th grade. We define elementary schools as treatment group and elementary-middle schools as control group.

Finally, all the variables we use in the empirical exercise are from the aforementioned two datasets (SAEB and Scholar Census), except for the per capita GDP, which is provided by the Brazilian Institute of Geography and Statistics (IBGE).

4. Methodology

Let Y_{1i} be the outcome of school i if it is treated (elementary school) and Y_{0i} the outcome of that school if it is not exposed to the treatment (elementary-middle school). The average treatment effect (ATE) for a particular school i can thus be written as

$$\mathbb{E}\big[Y_{1i} - Y_{0i}\big] \tag{5}$$

However, as it is impossible to assign the same school to both conditions, an alternative form of evaluation is the average treatment effect on treated (ATT). Let $Z \in \{0, 1\}$ be an indicator variable denoting the treatment received ($Z = 1$ is when the school is treated and $Z = 0$ when it is not). Thus, we denote ATT by

$$\mathbb{E}\big[Y_{1i} - Y_{0i} \,|\, Z_i = 1\big]. \tag{6}$$

The ATT is the mean effect of treatment for those schools which actually are only elementary schools. For policy purposes, the ATT is of interest rather than the ATE (Heckman, Ichimura, & Todd, 1997). The problem arises because we can observe only $Y_{1i}|Z_i = 1$ or $Y_{0i}|Z_i = 0$. The counterfactual $Y_{0i}|Z_i = 1$ is not observed and $\mathbb{E}\big[Y_{0i} \,|\, Z_i = 1\big] - \mathbb{E}\big[Y_{0i} \,|\, Z_i = 0\big]$ will probably be different to zero in non-experimental studies because the covariates that determine the treatment decision also determine the outcomes of interest. Therefore, differences between control and treatment groups remain even in the absence of treatment (self-selection bias). For our purpose, for example, richer municipalities may be more likely to maintain different schools for different school stages and such municipalities have more resources to invest in education.

In response to that, we must rely on some identifying assumptions to solve the selection problem. Rosenbaum and Rubin (1983) defined the following assumptions to ensure that a treatment assignment will be strongly ignorable:

$$\text{Unconfoundness} \qquad (Y_{1i}; Y_{oi}) \perp Z_i \,|\, X_i \,, \tag{7}$$

$$\text{Overlap} \qquad 0 < \Pr(Z_i = 1) \,|\, X_i < 1. \tag{8}$$

The first assumption says that in a set of observable covariates X which determines the treatment assignment, potential outcomes are independent of treatment status. The second condition ensures that the probability of being both participant and non-participant is bound away from zero and one. When both *unconfoundness* and *overlap* are held, then the following also holds:

$$\text{Unconfoundness given PS} \qquad (Y_{1i}; Y_{oi}) \perp Z_i \,|\, P(X_i), \qquad (9)$$

where the propensity score $P(X_i)$ is the probability of treatment given the school characteristics X. If the conditions for the strongly ignorable treatment assignment are held we can use the Propensity Score Matching (PSM) estimator for the ATT as follows:

$$ATT_{\text{PSM}} = \mathbb{E}_{P(x)|Z=1} \left\{ \mathbb{E}\left[Y_i \,|\, Z = 1, P(X) \right] - \mathbb{E}\left[Y_0 \,|\, Z = 0, P(X) \right] \right\}, \qquad (10)$$

that is, the PSM estimator is the difference between the mean outcome for elementary schools and the mean outcome for elementary-middle schools, weighted by the propensity score distribution. For this analysis, we use a logit model to estimate the probability of each school being only primary schools given X.

We use four different school outcomes as dependent variable: 5th grade students' average proficiency in Portuguese language and mathematics, average passing rate in elementary school and average dropout rate in elementary school. To select the covariates of the PSM we follow the procedure proposed by Imbens (2015), which allows, in addition to linear terms, second-order terms through an interactive process. All the first-order variables included in the estimation are shown in Table 4 that also shows the test of difference between treated and control groups.[5] The variables that represent characteristics of schools were extracted from the socioeconomic questionnaire of the Basic Education Assessment System (SAEB) while the municipal variable (*per capita* GDP) was taken from the IBGE. The PSM result also includes interactions, such as between administrative dependence and the municipality's GDP, which should avoid selection bias of schools in richer municipalities.

To verify the robustness of the results, in addition to the PSM estimation with the data from 2015, we conducted the same estimation with data from 2013 and an alternative methodology combining PSM with difference-in-difference between those years. We also ran tests to check the matching quality.

5. Results

In this section, we present the results obtained by the PSM methodology defined in section 4. Table A-1 in Appendix A shows the comparative statistics between

---

[5]The procedure implemented using the methodology proposed by Imbens (2015) removed only one of the covariates from our estimation, the principals average income.

elementary schools and elementary-middle schools. The descriptive statistics show that, even without any sort of controls, elementary schools have higher outcomes with better student characteristics.

In Table 1 we present the PSM estimates. The main methodology used to verify the outcomes is based on Kernel matching.[6] Following the approach used in Felício, Terra, and Zoghbi (2012) we also consider alternative PSM methodologies to check the results consistency.

Analyzing the four different outcomes, we can verify that the differences between the treatment and control groups, in our main methodology, are all statistically significant at 1%. The first column shows that the mean score in Portuguese Language is 4.25 points higher among students enrolled in elementary schools compared to those who are enrolled in elementary-middle schools. The results are similar for alternative methodologies (estimates range from 2.96 to 4.58).

Table 1. PSM estimates of the ATT for different outcomes, Brazil, 2015.

	Score of 5th-graders in Portuguese Language	Score of 5th-graders on Mathematics	Average passing rate in elementary school	Average dropout rate in elementary school
	(1)	(2)	(3)	(4)
Epanechnikov Kernel (bandwidth = 0.06)	4.25 *** (0.55)	5.09 *** (0.56)	2.46 *** (0.19)	−0.52 *** (0.05)
Nearest Neighbor without replacement	3.19 *** (0.25)	3.57 ** (0.27)	1.30 *** (0.09)	−0.15 *** (0.03)
Nearest Neighbor with replacement	3.09 * (1.62)	4.01 ** (1.67)	2.31 *** (0.60)	−0.36 ** (0.17)
Nearest 10 Neighbors with replacement	3.19 *** (0.84)	3.78 *** (0.85)	2.16 *** (0.30)	−0.42 *** (0.08)
Caliper 0.1	4.58 *** (0.45)	5.46 *** (0.46)	2.50 *** (0.16)	−0.50 *** (0.04)
Caliper 0.01	3.31 (0.85)	3.97 *** (0.87)	2.19 *** (0.30)	−0.43 *** (0.08)
Caliper 0.001	3.38 *** (0.60)	3.96 *** (0.61)	2.30 *** (0.21)	−0.42 *** (0.06)
Caliper 0.0001	2.96 *** (0.44)	3.41 *** (0.46)	2.09 *** (0.16)	−0.39 *** (0.04)

Notes: 29,963 observations. Standard errors in parentheses.   ***significant at 1%; **significant at 5%; *significant at 10%.

---

[6]Following Abadie and Imbens (2006), the standard errors of these estimates were not calculated by bootstrapping, because these values would not be valid while using Nearest Neighbor Matching.

Higher achievements among the treated group can also be verified in mathematics exams. Elementary schools students had their mathematics average score increased in 5.09 points. Considering the alternative methodologies, these estimates range from 3.41 to 5.46. As have seen, 5th grade primary schools students perform better in Portuguese language and in mathematics in average than those in elementary-middle school. The dimensions of these differences are in the order of 0.21–0.22 standard deviations.

As it can be seen in column 3, treatment group students have better chances to go on to the 6th grade as well. These results demonstrate that the elementary school had a positive and significant impact on the average passing rate of 5th graders of 2.46 percentage points (estimates range from 1.3 to 2.5). The size of the effect is considerably greater than found in achievements in Portuguese and mathematics, around 0.35 standard deviations.

The last column of Table 1 presents the ATT of dropout rate comparison (after matching) between 5th grade students, who are enrolled in elementary schools, and those who are enrolled in elementary-middle schools. The PSM estimates indicate that participating in the treated groups reduces the dropout rate in 0.52 percentage points (estimates range from 0.15 to 0.52). The size of these differences is in the order of 0.25 standard deviations.

Therefore, 5th grade students enrolled in elementary schools, when compared (after matching) to those from elementary-middle schools, present better school outcomes. The results maintain the same direction and significance for different specifications of the models.

5.1 Mechanisms

This subsection investigates possible variables that could be the channels that explain the best outcomes for students in the 5th year of elementary schools compared to elementary-middle schools. For example, the grade span configuration can lead to alternative pedagogical practices, focused on the specific needs of students. If this is a mechanism, the principals of these schools would tend to propose policies aimed at school learning, increase of approval rates and reduction of dropout. Another channel that can explain, at least part of the findings, is school management. Smaller schools may provide principals with more expertise in conducting school management.

The first groups of channels tested were those related to pedagogical practices. They are dummy variables assuming value 1 when the school offers specific policies for: achievement reinforcement, approval enhancement, dropout reduction and extracurricular activities. The second group of channels, related to school management, is formed by dummy variables when the school has: at least 90% of teachers with tenure, teacher training programs, and school board meetings at least quarterly and the continuous variable of the school average class size. These variables are

present in the SAEB database and are answered by school principals. These tests are performed using the same PSM regressions presented in Table 1, only replacing the dependent variables (school outcomes) by the dummy variables of the existence of alternative pedagogical practices. The results of the channel tests are shown in tables 2 and 3.

The results presented in Table 2 suggest that elementary schools are more likely to carry out alternative pedagogical policies. This result is not robust only for an achievement reinforcement policy. Holding the identification hypothesis of the PSM, we can say that primary schools are more likely, by 11 percentage points, to offer approval enhancement policies. Also, they are more likely to offer dropout reduction policies (25 percentage points) and extracurricular activities (5 percentage points). The findings for these three mechanisms are similar for different specifications. Regarding the test of the group of channels related to school management, shown in Table 3, the results also indicate that specialized schools tend to have benefits for teachers. There is no evidence, however, that these schools are more likely to

Table 2. PSM estimates of the ATT for the first channel group, Brazil, 2015.

	Scholar achievement reinforcement policy	Approval enhancement policy	Dropout reduction policy	Extracurricular activity policy
	(1)	(2)	(3)	(4)
Epanechnikov Kernel (bandwidth = 0.06)	0.01 (0.01)	0.11 *** (0.01)	0.25 *** (0.01)	0.05 *** (0.01)
Nearest Neighbor without replacement	0.03 *** (0.01)	0.13 *** (0.01)	0.28 *** (0.01)	0.04 *** (0.01)
Nearest Neighbor with replacement	−0.01 (0.02)	0.07 ** (0.04)	0.28 *** (0.04)	0.06 * (0.03)
Nearest 10 Neighbors with replacement	0.00 (0.01)	0.11 *** (0.02)	0.24 *** (0.02)	0.06 *** (0.02)
Caliper 0.1	0.01 * (0.01)	0.11 *** (0.01)	0.25 *** (0.01)	0.05 *** (0.01)
Caliper 0.01	0.00 (0.01)	0.11 *** (0.02)	0.23 *** (0.02)	0.06 *** (0.02)
Caliper 0.001	0.00 (0.01)	0.11 *** (0.01)	0.25 *** (0.01)	0.05 *** (0.01)
Caliper 0.0001	0.01 * (0.01)	0.11 *** (0.01)	0.26 *** (0.01)	0.04 *** (0.01)

Notes: 29,963 observations. Standard errors in parentheses.   ***significant at 1%; **significant at 5%; *significant at 10%.

Table 3. PSM estimates of the ATT for the second channel group, Brazil, 2015.

	At least 90% of teachers with tenure	Teacher training programs	School board meeting at least quarterly	Average class size
	(1)	(2)	(3)	(4)
Epanechnikov Kernel (bandwidth = 0.06)	0.10 *** (0.01)	0.05 *** (0.01)	−0.02 (0.01)	−1.39 *** (0,15)
Nearest Neighbor without replacement	0.15 *** (0.01)	0.04 *** (0.01)	−0.05 (0.01)	−0,08 (0,07)
Nearest Neighbor with replacement	0.13 *** (0.03)	0.06 * (0.03)	−0.05 (0.03)	−1.30 *** (0.43)
Nearest 10 Neighbors with replacement	0.09 *** (0.02)	0.06 *** (0.02)	−0.03 (0.02)	−1.47 *** (0.23)

Notes: 29,963 observations. Standard errors in parentheses.   ***significant at 1%; **significant at 5%; *significant at 10%.

have board meetings more often. The results indicate that elementary schools are more likely (10 percentage points) to have at least 90% of their teachers with tenure and probably (5 percentage points) offer their teachers training programs. Schools in the treated group also have a smaller class size in approximately 1.3 students on average. These findings suggest that both mechanisms (pedagogical practice and school management policies for teachers) may be acting to explain the impacts of grade span configuration in the school outcomes of 5th grade students.

6. Testing

6.1 Matching quality

For the purpose of checking the quality of PSM, we test if the covariates used in the matching process were properly balanced between treatment and control groups. The first test, proposed by Rosenbaum and Rubin (1985), consist of a *t* test to check the differences between the averages of each covariate of treatment and control groups. The results shown in Table 4 are based on Kernel PSM, which was chosen as the main methodology of this paper.

The first two columns of Table 4 show the mean value of each covariate for treated and control groups, respectively. The third column, shows the t statistic of the differences between the groups after matching. The *t*-statistic, shown in the third column, shows that all the *p*-values for these differences are higher than 10%, indicating that the matching was adequately balanced. This means that the difference of the covariates between the treatment and control groups are not statistically significant at the 10% threshold.

To support the theory that the matching was properly balanced, we also consider Kernel density functions before and after matching to show the overlap of these distributions. Figure 2 shows density functions that overlap almost perfectly after matching, indicating a precise balance of the PSM.

Table 4. Variable Analysis.

|	Treated	Control	t-statistic after balancing
Dummy representing the location of the school	0.913	0.914	0.646
Dummy representing the administration of the school	0.864	0.862	0.563
Dummy representing the choice of the principal	0.482	0.465	0.001
Percentage of principals with a postgraduate degree	0.828	0.807	0
Percentage of teachers with a postgraduate degree	0.045	0.052	0
Mean years of experience of the principals	4.574	4.794	0
Mean years of experience of the teachers	13.979	14.125	0.006
Mean teachers' income	2,151	2,135	0.148
Dummy of Socioeconomic Level 1	0.003	0.003	0.951
Dummy of Socioeconomic Level 2	0.042	0.042	0.819
Dummy of Socioeconomic Level 4	0.284	0.281	0.586
Dummy of Socioeconomic Level 5	0.383	0.386	0.672
Dummy of Socioeconomic Level 6	0.112	0.112	0.897
Dummy of Socioeconomic Level 7	0.001	0.001	0.177
Per capita GDP	25,908	26,497	0.009
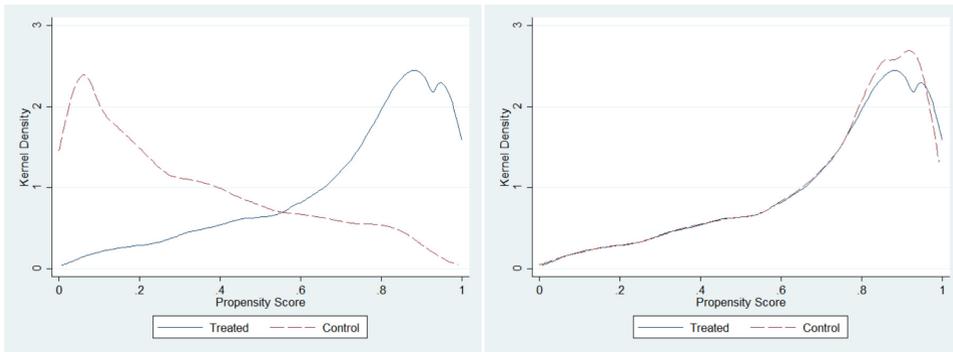
Figure 2. Kernel density functions before and after matching.

6.2 Sensitivity analysis of the results of unobserved heterogeneity

Even though many controls are used to assure the quality of the PSM, if both control and treatment groups differ on unobserved variables that may affect treatment, or outcome, there is a possibility of the existence of a "hidden bias". This problem, with non-experimental data, cannot be solved since there is no way to measure the effect

of this bias against the outcome variables. In order to check the existence of hidden biases, we used the implemented solution suggested by Rosenbaum (2002) that determines how much an unmeasured variable influences the treatment assignment or the outcome.

This issue with non-experimental analysis can also be found in Aakvik (2001); Caliendo, Hujer, and Thomsen (2005); and Caliendo, Hujer, and Thomsen (2008). This implementation can be overall explained as the participation probability of the individual i with the observed characteristics X_i. If the equation that describes this probability includes all the important variables, and there is no hidden bias, we can conclude that γ will be zero and the participation in the treatment will only be determined by the explicit covariates. Following Aakvik (2001) and Caliendo et al. (2008), for the significance of effects, we gradually increased the e^γ until the inference about the treatment effect changed. With this we can analyze the strength that unmeasured covariates would require to change the intuition about the treatment effect.

The results shown in Table 5 check if the individuals have the same probability of receiving treatment, and this happens when $e^\gamma = 1$, therefore $\gamma = 0$. If, the critical value for $e^\gamma = 2$, individuals who appear to be similar, may differ in their probability of receiving the treatment by a factor of 2. In other words, e^γ is a measure of the degree that the propensity is free of hidden bias. These values do not represent that the model contains unobserved bias. It simply verifies that, in case of the existence of hidden biases, the confidence interval, until the inference was changed, is contained in these values. The results allow us to conclude that even large amounts of unobserved heterogeneity would not alter the inference about the ATT shown in Table 1.

Table 5. Sensitivity Analysis for Unobserved Heterogeneity – Critical value for e^γ.

	Score of 5th-graders on Portuguese Language	Score of 5th-graders on Math	Average passing rate in elementary school	Average passing rate in elementary school
	(1)	(2)	(3)	(4)
Epanechnikov Kernel (bandwidth = 0.06)	1.4–1.45	1.35–1.4	2.3–2.35	> 3
Nearest Neighbor without replacement	1.05–1.1	1.05–1.1	1.1–1.15	1
Nearest Neighbor with replacement	1.1–1.15	1.15–1.2	1.45–1.5	> 3
Nearest 10 Neighbors with replacement	1.2–1.25	1.15–1.2	2–2.05	> 3
Caliper 0.1	1.45–1.5	1.4–1.45	2.35–2.4	> 3
Caliper 0.01	1.25–1.3	1.2–1.25	2.1–2.15	> 3
Caliper 0.001	1.2–1.25	1.2–1.25	2–2.05	> 3
Caliper 0.0001	1.05–1.1	1.05–1.1	1.45–1.5	> 3

6.3 Alternative Identification Strategies

In order to check the robustness of the findings, we also ran PSM for the SAEB 2013, the previous year. As seen in Table 6, we can conclude that the results can be seen throughout the years and do not qualify as singular occurrences.

Another estimation strategy used to check whether the results are consistent is the PSM with difference-in-differences between 2013 and 2015. Our treated group are schools that have become elementary schools between 2013 and 2015. The control group are schools that remained elementary-middle schools between the two periods. Although not so strong, the results shown in Table 7 are similar to those found in tables 1 and 6. Note that in this specification the sample size is reduced. This is because the matching is done only with schools that have changed the grade configuration. Another reason that may explain the reduction of the magnitude of

Table 6. PSM estimates of the ATT for different outcomes, Brazil, 2013.

	Score of 5th-graders in Portuguese Language	Score of 5th-graders in Mathematics	Average passing rate in elementary school	Average dropout rate in elementary school
	(1)	(2)	(3)	(4)
Epanechnikov Kernel (bandwidth = 0.06)	6.47 *** (0.64)	7.45 *** (0.72)	2.63 *** (0.19)	−0.57 *** (0.06)
Nearest Neighbor without replacement	3.58 *** (0.31)	4.42 *** (0.35)	1.00 *** (0.09)	−0.18 *** (0.03)
Nearest Neighbor with replacement	3.95 ** (1.66)	5.31 *** (1.88)	2.14 *** (0.53)	−0.42 ** (0.17)
Nearest 10 Neighbors with replacement	5.47 *** (0.99)	6.51 *** (1.12)	2.60 *** (0.31)	−0.51 *** (0.10)
Caliper 0.1	6.64 *** (0.52)	7.69 *** (0.60)	2.58 *** (0.16)	−0.56 *** (0.05)
Caliper 0.01	4.83 *** (1.18)	5.75 *** (1.33)	2.43 *** (0.35)	−0.48 *** (0.11)
Caliper 0.001	4.76 *** (0.83)	5.59 *** (0.94)	2.34 *** (0.25)	−0.44 *** (0.08)
Caliper 0.0001	4.79 *** (0.53)	5.48 *** (0.60)	2.15 *** (0.17)	−0.41 *** (0.05)

Notes: 29,922 observations. Standard errors in parentheses. ***significant at 1%; **significant at 5%; *significant at 10%.

Table 7. Diff-in-diff estimates of the ATT for different outcomes, Brazil, 2013.

	Score of 5th-graders in Portuguese Language	Score of 5th-graders in Mathematics	Average passing rate in elementary school	Average dropout rate in elementary school
	(1)	(2)	(3)	(4)
Epanechnikov Kernel (bandwidth = 0.06)	2.73 **	3.46 ***	2.28 ***	−0.67 ***
	(1.09)	(1.30)	(0.43)	(0.22)
Nearest Neighbor without replacement	0.82	0.57	2.34 ***	−0.66 **
	(1.59)	(1.94)	(0.65)	(0.28)
Nearest Neighbor with replacement	−6.93	−6.80 *	1.63 **	−0.59
	(4.27)	(3.79)	(1.63)	(0.62)
Nearest 10 Neighbors with replacement	2.23	3.63 *	1.42 **	−0.23
	(1.75)	(2.09)	(0.70)	(0.27)
Caliper 0.1	2.88 ***	3.76 ***	2.22 ***	−0.67 ***
	(1.10)	(1.29)	(0.43)	(0.22)
Caliper 0.01	1.82 *	2.17 *	2.14 ***	−0.68 ***
	(1.09)	(1.30)	(0.44)	(0.22)
Caliper 0.001	1.39	1.50	2.03 ***	−0.65 ***
	(1.09)	(1.30)	(0.44)	(0.22)
Caliper 0.0001	1.40	1.53	2.03 ***	0.66 ***
	(1.09)	(1.30)	(0.44)	(0.22)

Notes: 367 observations. Standard errors in parentheses. ***significant at 1%; **significant at 5%; *significant at 10%.

the coefficient is that the effects may be lower due to the short time of exposure to the treatment. Changing pedagogical and management practices may take some time.

7. Conclusion

The main purpose of this paper was to analyze, given the demographic transition that Brazil is going through, the alternative courses of action given the excess supply in the Brazilian school system. As seen in other countries, this demographic change creates an opportunity to reformulate the educational system. This could be done by closing schools and relocating students, or by specializing schools, splitting the elementary-middle school configuration into elementary and middle schools, in separate sites. This phenomenon was already studied in the United States (Byrnes & Ruby, 2007). Their comparison was among students that underwent a school change in their lives, and how it affected their proficiency outcomes. Even though

their results show that this school transition affects the performance of students in standardized tests negatively, their focus was to observe this effect in older students, entering high-school. Our objective was to analyze if this school division, into elementary and middle school, could positively affect the performance of 5th grade students in Brazil. Our analysis is based on a dataset from the Brazilian Ministry of Education, which provides information on both students' standardized test results and socioeconomic information of students, teachers and principals.

We estimated the effects of grade span configuration on different outcomes, using the Propensity Score Matching strategy. There are four preliminary findings. First, treated group of school students presented results on standardized tests of Portuguese language (0.21 standard deviations) higher than those achieved by the control group. Second, students from these schools also present higher achievement in mathematics (0.22 standard deviations). Third, we estimate that the average passing rate is 2.46 percentage points higher in elementary schools compared to elementary-middle schools. This result represents a positive impact of 0.35 standard deviations. Fourth, there is evidence that these schools are also successful in reducing the dropout rate by 0.52 percentage points (0.25 standard deviations). We found evidence that suggests that both alternative pedagogical practices and school management policies for teachers are important mechanisms of the effects of school specialization on student outcomes.

The results presented in this study are statistically significant and robust to different methodological specifications. Therefore, we believe there is evidence that a change in grade span configuration, which specializes schools, can improve the school outcomes of 5th grade students. These findings have interesting implications for the implementation of public policies. A reorganization of the school structure is favorable in an environment of reduction of enrollments and maintenance of the schools supply.

Finally, this study has the limitation of only analyzing the achievements of 5th grade students. Other studies may contribute to the literature evaluating possible effects for students who switch schools with the change in grade span configuration.

References

Aakvik, A. (2001). Bounding a matching estimator: The case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics*, *63*(1), 115–143. http://dx.doi.org/10.1111/1468-0084.00211

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267. http://dx.doi.org/10.1111/j.1468-0262.2006.00655.x

Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics*, *22*(2), 141–165. http://dx.doi.org/10.1080/09645292.2011.568698

Byrnes, V., & Ruby, A. (2007). Comparing achievement between k–8 and middle schools: A large-scale empirical study. *American Journal of Education*, *114*(1), 101–135. http://dx.doi.org/10.1086/520693

Caliendo, M., Hujer, R., & Thomsen, S. L. (2005). *Individual employment effects of Job Creation Schemes in Germany with respect to sectoral heterogeneity* (Discussion Paper No. 13/2005). Nürnberg. http://doku.iab.de/discussionpapers/2005/dp1305.pdf

Caliendo, M., Hujer, R., & Thomsen, S. L. (2008). The employment effects of job-creation schemes in Germany: A microeconometric evaluation. In T. Fomby, R. C. Hill, D. L. Millimet, J. A. Smith, & E. J. Vytlacil (Eds.), *Modelling and evaluating treatment effects in econometrics* (pp. 381–428). Emerald Group Publishing. (Advances in Econometrics, Vol. 21) http://dx.doi.org/10.1016/S0731-9053(07)00013-8

Dhuey, E. (2013). Middle school or junior high? How grade-level configurations affect academic achievement. *Canadian Journal of Economics*, *46*(2), 469–496. http://dx.doi.org/10.1111/caje.12020

Dove, M. J., Pearson, L. C., & Hooper, H. (2010). Relationship between grade span configuration and academic achievement. *Journal of Advanced Academics*, *21*(2), 272–298. http://dx.doi.org/10.1177/1932202X1002100205

Felício, F. d., Terra, R., & Zoghbi, A. C. (2012). The effects of early childhood education on literacy scores using data from a new Brazilian assessment tool. *Estudos Econômicos*, *42*(1), 97–128. http://dx.doi.org/10.1590/S0101-41612012000100004

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

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*(5), 449–481. http://dx.doi.org/10.1016/S0883-0355(03)00015-6

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, *64*(4), 605–654. http://dx.doi.org/10.2307/2971733

Holmlund, H., & Böhlmark, A. (2019). Does grade configuration matter? Effects of school reorganisation on pupils' educational experience. *Journal of Urban Economics*, *109*, 14–26. http://dx.doi.org/10.1016/j.jue.2018.11.004

Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, *50*(2), 373–419. http://dx.doi.org/10.3368/jhr.50.2.373

Offenberg, R. M. (2001). The efficacy of philadelphia's k-to-8 schools compared to middle grades schools. *Middle School Journal*, *32*(4), 23–29. http://dx.doi.org/10.1080/00940771.2001.11495283

Rockoff, J. E., & Lockwood, B. B. (2010). Stuck in the middle: Impacts of grade configuration in public schools. *Journal of Public Economics*, *94*(11), 1051–1061. http://dx.doi.org/10.1016/j.jpubeco.2010.06.017

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4757-3692-2

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. http://dx.doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38. http://dx.doi.org/10.2307/2683903

Schanzenbach, D. W. (2014). *Does class size matter?* Boulder, CO: National Education Policy Centre (NEPC). https://nepc.colorado.edu/publication/does-class-size-matter

Schwerdt, G., & West, M. R. (2013). The impact of alternative grade configurations on student outcomes through middle and high school. *Journal of Public Economics*, *97*, 308–326. http://dx.doi.org/10.1016/j.jpubeco.2012.10.002

Weiss, C. C., & Bearman, P. S. (2007). Fresh starts: Reinvestigating the effects of the transition to high school on student outcomes. *American Journal of Education*, *113*(3), 395–421. http://dx.doi.org/10.1086/512738

Appendix A. Omitted proofs and Table A-1

Proof of Proposition 1

Let us start by proving the existence of solution. Define the function $H : [0, \infty) \to \mathbb{R}$ by $H(y_j) = 2p'(y_j)(\mu_j - p(y_j)) - C_y(y_j, x)$ and notice that

$$H(0) = 2p'(0)(\mu_j - p(0)) > 0, \tag{A-1}$$

$$H(\hat{y}_j) = -C_y(\hat{y}_j, x) < 0, \tag{A-2}$$

where we use the properties of $p(\cdot)$ and $C(\cdot)$. Because $H(\cdot)$ is a continuous function, the intermediate value theorem applies, and thus there exists $y_j(x) \in (0, \hat{y}_j)$ such that $H(\hat{y}_j(x)) = 0$.

Uniqueness can be proven through the second order condition (SOC). Formally,

$$H_y(y_j) = 2\left[p''(y_j)(\mu - p(y_j)) - (p'(y_j))^2\right] - C_{yy}(y_j, x) < 0, \tag{A-3}$$

where we once again use the properties of $p(\cdot)$ and $C(\cdot)$.

Observe now that by taking the implicit derivative of (3) we have

$$y_j'(x) = -\frac{-C_{yx}(\hat{y}_j, x)}{H_y(y_j)} > 0. \tag{A-4}$$

By using the properties of $C(\cdot)$, it is straightforward to show that $\lim_{x \to +\infty} y_j'(x) = 0$. Finally, recalling that $0 < y_j(x) < \hat{y}_j$, we conclude that when $x \to \infty$ we have $y_j(x) \to \hat{y}_j$ and therefore $y_j''(x) < 0$. $\qquad\square$

Proof of Proposition 2

The strategy of proof is similar to the previous proposition. We start by establishing existence. Define

$$G(x) = 2\,\mathbb{E}\left\{p'(y_j(x; \mu_j))y_j'(x; \mu_j)[\mu - p(y_j(x; \mu_j))]\right\} - B_x(x; \sigma_*)$$

$$= 2\int p'(y_j(x; \mu_j))y_j'(x; \mu_j)[\mu - p(y_j(x; \mu_j))]\,dF(\mu) - B_x(x; \sigma_*), \tag{A-5}$$

then observe that

$$G(0) = 2\int p'(y_j(x; \mu_j))y_j'(x; \mu_j)[\mu - p(y_j(x; \mu_j))]\,dF(\mu) > 0 \tag{A-6}$$

$$\lim_{x \to \infty} G(x) = -\infty, \tag{A-7}$$

because $p'(y_j(x; \mu_j))y_j'(x; \mu_j)[\mu - p(y_j(x; \mu_j))] \to 0$ and $B_x(x; \sigma_*) \to +\infty$. Given that $G(\cdot)$ is continuous, the intermediate value theorem applies and thus there exists $x^* \in (0, \infty)$ such that $G(x^*) = 0$.

Taking the first derivative of $G(\cdot)$ we have the SOC of the principal's problem:

$$2\int\left\{\left[p''\big(y_j(x;\mu_j)\big)\big(y_j'(x;\mu_j)\big)^2 + p'\big(y_j(x;\mu_j)\big)y_j''(x;\mu_j)\right]\left[\mu - p\big(y_j(x;\mu_j)\big)\right]\right.$$

$$\left. - 2\big(p'\big(y_j(x;\mu_j)\big)\big)y_j'\big(x;\mu_j\big)\big)^2 dF(\mu)\right\} - B_{xx}(x;\sigma_*) < 0, \quad \text{(A-8)}$$

where we use the properties of functions $p(\cdot)$ and $B(\cdot)$, and the result os Proposition 1. This proves the strictly concavity of the principal's utility function and then the uniqueness of the solution.   □

Proof of Proposition 3

First, observe that

$$S = \int_{p_j-\Delta}^{p_j+\Delta} d\Phi(\mu) \tag{A-9}$$

$$= \Phi\big(p\big(y_j(x^*)\big)+\Delta\big) - \Phi\big(p\big(y_j(x^*)\big)-\Delta\big). \tag{A-10}$$

We are interested in the signal of $dS/d\sigma_*$, which can be obtained through the chain rule:

$$\frac{dS}{d\sigma_*} = \frac{dS}{dy_j}\frac{dy_j}{dx}\frac{dx^*}{d\sigma_*}. \tag{A-11}$$

We have already showed that $y_j'(x) > 0$. Now, we can apply the implicit function theorem in (4) and thus obtain

$$\frac{dx^*}{d\sigma_*} = -\frac{-B_{x\sigma}(x;\sigma_*)}{G_x} < 0, \tag{A-12}$$

because $G_x > 0$, as we showed in Proposition 1, and $B_{x\sigma}(x;\sigma_*) > 0$ by assumption. Finally, notice that

$$\frac{dS}{dy_j} = \left[\phi\big(p\big(y_j(x^*)\big)+\Delta\big) - \phi\big(p\big(y_j(x^*)\big)-\Delta\big)\right]p'\big(y_j(x^*)\big) > 0, \tag{A-13}$$

given that $p'\big(y_j(x^*)\big) > 0$, and because $\phi(\cdot)$ is the f.d.p. of the normal distribution, we have $\phi\big(p\big(y_j(x^*)\big)+\Delta\big) > \phi\big(p\big(y_j(x^*)\big)-\Delta\big)$ for any $p\big(y_j(x^*)\big) < \mu^*$, which is always the case, as we have showed. Therefore, we can use the chain rule in (A-11) to conclude that $dS/d\sigma_* < 0$.   □

Table A-1. Descriptive statistics.

	Elementary Schools					Elementary-middle Schools				
	Obs (% within category)	Mean Score in Portuguese Language (Std.Dev)	Mean Score in Mathematics (Std.Dev)	Passing Rate (Std.Dev)	Dropout rate (Std.Dev)	Obs (% within category)	Mean Score in Portuguese Language (Std.Dev)	Mean Score in Mathematics (Std.Dev)	Passing Rate (Std.Dev)	Dropout rate (Std.Dev)
All Schools	16317 (53.86)	210.35 (20.10)	223.45 (22.49)	93.07 (6.78)	0.90 (2.03)	13979 (46.14)	205.60 (21.94)	217.88 (22.28)	91.55 (7.71)	1.16 (2.14)
State Schools	2214 (7.31)	217.54 (17.74)	230.78 (20.51)	95.85 (5.44)	0.63 (1.67)	4033 (13.31)	214.36 (17.13)	226.12 (17.85)	94.29 (6.98)	0.66 (1.87)
City Schools	14103 (46.55)	209.22 (20.21)	222.30 (22.58)	92.63 (6.87)	0.95 (2.08)	9946 (32.83)	202.05 (22.68)	214.55 (23.01)	90.44 (7.71)	1.36 (2.20)
Urban Schools	14876 (49.10)	211.93 (19.19)	224.94 (21.84)	93.38 (6.56)	0.87 (2.03)	11051 (36.48)	210.33 (19.15)	222.11 (20.15)	92.26 (7.30)	1.03 (1.99)
Country Schools	1441 (4.76)	194.00 (21.88)	207.99 (23.36)	89.90 (8.15)	1.26 (2.05)	2928 (9.66)	187.75 (22.63)	201.94 (22.69)	88.88 (8.57)	1.65 (2.54)
Socioeconomic Level 1	61 (0.20)	172.68 (21.26)	188.78 (21.24)	85.12 (10.13)	2.03 (2.35)	202 (0.67)	163.81 (14.20)	180.51 (11.60)	84.23 (10.21)	2.39 (2.95)
Socioeconomic Level 2	691 (2.28)	182.13 (18.23)	196.49 (20.17)	87.13 (8.84)	2.26 (2.97)	1304 (4.30)	181.87 (23.56)	197.00 (24.64)	88.02 (9.02)	2.02 (2.73)
Socioeconomic Level 3	2852 (9.41)	190.33 (16.12)	202.36 (17.28)	88.92 (7.94)	1.90 (2.63)	2905 (9.59)	190.12 (18.71)	203.16 (19.79)	88.69 (8.44)	2.11 (2.82)
Socioeconomic Level 4	4629 (15.28)	204.89 (14.23)	216.52 (16.79)	92.20 (6.67)	1.11 (2.05)	3072 (10.14)	203.51 (14.37)	214.74 (15.76)	91.16 (7.87)	1.28 (2.22)
Socioeconomic Level 5	6240 (20.60)	219.70 (11.65)	233.67 (15.00)	95.14 (4.91)	0.37 (1.51)	4701 (15.52)	215.37 (12.14)	227.09 (13.68)	93.38 (6.25)	0.56 (1.15)
Socioeconomic Level 6	1831 (6.04)	235.07 (10.52)	249.99 (13.92)	97.16 (2.92)	0.10 (0.36)	1767 (5.83)	230.36 (11.93)	242.11 (13.92)	95.39 (3.99)	0.20 (0.57)
Socioeconomic Level 7	13 (0.04)	251.09 (10.00)	266.78 (14.44)	98.45 (1.55)	0.00 (0.00)	28 (0.09)	244.70 (8.83)	257.92 (14.20)	98.51 (1.54)	0.04 (0.19)
North	1814 (5.99)	195.75 (17.86)	206.38 (18.69)	90.53 (7.05)	1.57 (2.13)	1398 (4.61)	186.73 (20.02)	198.70 (18.33)	86.84 (8.61)	2.47 (2.91)
Northeast	4486 (14.81)	192.86 (16.82)	204.05 (16.97)	88.57 (7.75)	1.99 (2.79)	4218 (13.92)	189.84 (20.81)	202.87 (21.53)	88.23 (8.38)	2.08 (2.82)
Southeast	6522 (21.53)	221.41 (13.47)	236.09 (16.07)	96.02 (4.86)	0.33 (1.43)	4078 (13.46)	218.00 (14.03)	230.47 (15.58)	94.84 (6.07)	0.56 (1.16)
South	2396 (7.91)	221.94 (13.78)	237.86 (16.46)	94.93 (4.24)	0.14 (0.38)	3208 (10.59)	217.58 (14.63)	229.66 (15.18)	93.01 (5.74)	0.34 (0.73)
Central-West	1099 (3.63)	214.86 (13.62)	224.34 (14.80)	94.09 (4.86)	0.46 (0.84)	1077 (3.55)	209.16 (13.95)	218.90 (14.51)	93.90 (6.14)	0.54 (1.03)