São Paulo School of ECONOMICS

Working Paper

May 2011

CENTER FOR APPLIED MICROECONOMICS

10/2011

C-MICRO Series

The impact of daycare attendance on Math test scores for a cohort of 4th graders in Brazil

Clarissa G. Rodrigues

Cristine C. X. Pinto

Daniel D. Santos

DRAFT RESEARCH PAPER
Not for citation or circulation

The impact of daycare attendance on Math test scores for a cohort of 4^{th} graders in Brazil

by: Clarissa G. Rodrigues (CEDEPLAR), Cristine X.C. Pinto (EESP/FGV) and Daniel D. Santos (IBMEC/RJ)

Initial Draft: October 2009
This Draft: April 13, 2011

Abstract

We estimate the impact of having attended center-based daycare institutions during early childhood on Math test scores at the 4^{th} grade of elementary school. Because enrollment in daycare centers may depend on unobservable characteristics of the family and the child, we build and estimate a structural model of endogeneous choice of school to deal with the selectivity problem. We find that attendance to daycare institutions is associated with a gain of approximately 0,04 standard deviation in Math test scores. This result is important to the extent our OLS results as well as most of the studies for Brazil find no effect associated to daycare attendance, suggesting selectivity may play a role on this finding.

1 Introduction

Investments in Early Childhood Development (ECD) have increasingly being pointed as a key ingredient to understand the process of human capital formation. On one hand, there is neurological evidence that learning is easier during this stage and stimulation of cognitive abilities may also facilitate learning later in life[1]. On the other hand, economists have argued that, *ceteris paribus*, returns to investments made earlier in life are usually higher because individuals have a longer time to reap the benefits[2]. Moreover, if the abilities acquired earlier are inputs to future human capital accumulation, this is a second reason to believe that returns to ECD investments are probably higher than other investments in human capital[3].

The consensus researchers reached on the importance of ECD investments contrasts with prevailing divergence about the best way to promote it. People often agree that parental care is the best use of time of a newborn child as well as that no child should be left out of school after 6 or 7 years old, but there is intense debate about the benefits of time spent in kindergarten and/ or daycare institutions. The predominant evidence for the US suggests that large scale programs seem to have significant impacts on test scores taken at the beginning of elementary school, but tend to fade out over time. Magnusson et al (2004, 2007) for example find the impact of pre-kindergarten attendance on kindergarten performance in Math and language is around 0,15 standard deviation (being a little higher for children from vulnerable families), but declines to 0,03 s.d. by the spring of the first grade of elementary school. Currie at al. (1995) find similar results for the Head Start, a large-scale compensatory center-based educational program focused on vulnerable children. The authors find that about one third of the gap in PPVT scores[4] between Head Start attendees and their more advantaged peers is closed because of the participation in the program, but this impact declines over time, eventually vanishing for blacks at the age of 9[5]. Small-scale model programs such as the Chicago Child-Parent Centers[6], the High Scope/ Perry Preschool[7] and the Carolina Abecedarian Project[8], display high and persistent effects on high school conclusion rates, measures of cognitive abilities and labor outcomes, but their high cost and the difficulty in replicating the prevailing conditions at which they were implemented raise a

[1] See Borghans et al (2008) and Cunha and Heckman (2009) for a discussion).

[2] See Becker (1993).

[3] Carneiro and Heckman (2003).

[4] Image-based cognitive test applied to 3-4 year-old children.

[5] In a companion paper, Currie et al (2000) justify this result with the argument that, among former Head Start attendees, whites are usually enrolled in better schools than blacks, and it is this complementarity that makes the impact of Head Start to fade out among blacks (and to persist among whites).

[6] See an evaluation of the CPC program in Temple and Reynolds (2007), Temple et al (2000).

[7] Implemented in the mid-1960's in Ypsilanti, Michigan, and evaluated by Belfield et al (1993) and Berrueta-Clement et al (1984).

[8] Implemented in the early 1970's. See evaluations by Campbell et al (2001, 2002 and 2008).

concern about their validity as an instrument of policy[9].

The evidence for Latin America is far less conclusive. Using Argentinian data, Berlinski et al (2009) find that preschool attendance improves transcript grades of third graders by 0,23 s.d., whereas Berlinski et al (2008) find no impact of preschool attendance on elementary school enrollment in Uruguay, and a non-robust positive impact on the number of completed grades of 7-11 year-old children. Interestingly, the authors conclude that in the models the impact appear to be significant, it is the attendance to preschool that matters, and not the length of exposition to these services. Behrman et al (2004) find that participation in the Bolivian PIDI[10], a nutritional and educational center-based program for 0 to 72-month poor children, have a positive effect on language test scores for those exposed at least 7 months to the intervention.

We aim to contribute to this literature by estimating the average impact of daycare and preschool attendance on cognitive outcomes measured at the 4^{th} grade of elementary school. The key problem with this estimation is that children may not be sent to daycare and preschool institutions at random. If families and children are unobservationally heterogeneous, one may for example think that the best children are the ones that start education earlier (if innate talent is complementary with school outputs), in which case naive OLS regressions including indicators of attendance to daycare/ preschool would be upward biased. On the other hand, it may be the case that the worst children are the ones sent to center-based ECD institutions, if families see in this attitude a chance of compensating disadvantaged children for their handicap. In both cases, there is room for a potential problem of endogeneity in these indicators, and we intend to address this problem by using an empirical strategy that a theoretical model predicts would be appropriate to solve this problem. In this strategy, we use indicators of supply of daycare and preschool (kindergarten) facilities in the municipality, and of incidence of contagious diseases among children at preschool age as instruments that would shift the decision of sending individuals to daycare/ preschool institutions without directly influence test scores.

The rest of the paper continues as follows. In the next section we present a descriptive analysis of the relationship between daycare attendance in the early childhood and test scores in elementary and high school. Since enrollment in daycare institutions is not mandatory, some attention is given to the fact that families that choose to send their children to these centers may be different from those who do not send. If this is the case, differences in test scores between former daycare center attendees and non-attendees may be due to intrinsic differences in characteristics of these groups and not to the attendance itself. In fact, we show that children who attended daycare centers come from wealthier families and are enrolled in better elementary and high schools at the moment the test is applied.

In Section 3 we present a theoretical model of human capital formation

[9] See also Barnett (1992) and Blau and Currie (2006) for surveys of the results of other evaluations of ECD educational programs in US.

[10] Proyecto Integral de Desarrollo Infantil

with endogenous decisions about when to firstly send children to school and of which quality. We then proceed to estimate the effect of daycare attendance on test scores accounting for the potential endogeneity in the indicator of daycare attendance. Our database is composed by the results of the SAEB, a biennial test applied by the Brazilian ministry of education to a sample of students of the 4^{th}, 8^{th}, and 11^{th} grades, covering Math and language subjects. We merge these data with municipal-level information about contagious diseases among children, from the Unified Health System (SUS), and about the supply of daycare and preschool, from the Educational Census, and use variables related to the supply of vacancies in daycare centers and health as instruments for the decision of daycare attendance. The last section contains our final remarks.

2 Overview

The genesis of the Brazilian daycare system is intrinsically associated to the emancipation of women and their need to participate in the labor market. It was only in the 90's that the findings about the importance of skill formation during early childhood have reshaped the Brazilian daycare system towards more emphasis in educational aspects. In this period, not only the daycare system gained a fully structured curriculum to be followed between ages 1 to 3, but also has the law changed, by including access to daycare centers as a basic right of all children[11], and by reducing the minimum age for which attending school is mandatory, first from 7 to 6 years old[12] and later changed to 4 years old[13]. Together with such changes, the government (both at the national and local levels) has created incentives for the families to send their children to school at early ages, both by raising the funds directed to public daycare centers and kindergarten[14], by raising the minimum qualification recquired for being a teacher in a daycare/ kindergarten center[15], and by giving advice for the families.

In this study, we use data from the SAEB 2005, a national exam of Math and language taken by a random sample of students at the end of the elementary school (4th grade), junior high (8th grade), and high school (11th grade)[16]. The children in our database therefore belong to the first cohorts of individuals who attended the ECD educational system after the change in its guidance. Together with the tests, the children provide information about individual and family characteristics, including a question about the age they first went to school (the children has to indicate one out of four alternatives: (i) daycare centers, (ii) kindergarten, (iii) first grade of elementary school, (iv) second or higher grade

[11] Constitution of 1988, improved on the Statute of the Children and Teenagers (1990).

[12] National Educational Bases and Guidelines Law (LDB), of 1996.

[13] 2009 LDB ammendment.

[14] After the LDB of 1996, daycare centers become the only educational level for which private institutions may receive long term public transfers.

[15] College degree. The law establishe a 3-year period for existing schools to be integrated to the new regulation, but local legislators may give schools some extra time to do so.

[16] For a detailed description of SAEB, see Appendix 1).

of elementary school). We use this information to investigate whether early entrance has a significant impact on Math and language scores.

Table 1 below shows that in fact former daycare centers (DC) attendees perform, on average, 0.3 standard deviations above the overall mean, and around 0.7 sd above the mean of the children who entered only at the first grade of elementary school. This table also shows that the test score differences by the first school experience do not fade out throughout the educational cycle (even though the data refer to different cohorts in a system which is rapidly changing over time), and that these patterns are similar for Math and language[17].

Of course, the gross differentials displayed above do not necessarily capture the causal effect of early entrance in school on the formation of cognitive skills, as measured by test scores. If the timing of the first educational experience is correlated with other variables that determine test scores, then the differentials in Table 1 may be contaminated by the effect these variables have on test scores. Since families are free to choose the proper timing for sending their children to school, we expect they do this in an optimal way, which may vary across families and children with different characteristics and generate a sistematic correlation between the daycare decision and such characteristics. In this section we will act as if we were able to observe all of these characteristics, and therefore to control for all of these correlations in the measurement of the causal effect of daycare attendance on test scores. The next section will focus on the potential correlation with unobservables.

We identify at least four channels that should be accounted for in an exercise that aims to identify the net effect of daycare attendance on test achievement. First, it is possible that families are heterogeneous in the way they value human capital formation of their children. If the families that send children earlier to school are those who value school the most, then we should expect that children who attended daycare centers are those who also benefit from a more stimulating home environment, attend better schools in later stages of their formation, etc. If this is the case, the gross differential in test scores between former DC students and non-DC students may be due to differences in the overall family investment in education instead of the treatment efefct of daycare attendance itself.

Second, it is possible that parents with higher opportunity cost of time are those who send children to daycare centers, in order to be able to explore their advantages in the labor market. Since some characteristics that determine the market value of labor are also important in the formation of the child's human capital (e.g. parents' education), it is possible that part of the test score variation between DC and non-DC students is due to the fact that the parents of the first have more human capital than those of the later, and not to DC attendance itself.

Third, it is possible that parents with particularly low ability in child care-taking are those who choose to send their children to DC institutions. If this is the case, it may be the children of the "worst" families that go early to school,

[17] Students who took language exams at the 11th grade did not provide socioeconomic information, and were excluded from our analysis.

and part of the gross difference showed in Table 1 could be attributable to the parental inability of promoting child development.

Fourth, it is possible that sending children to daycare centers is a way of compensating deficiencies in child development. If this is the case, children with worst cognitive/ non-cognitive skills would more likely attend daycare centers.

From this discussion we conclude that properly controlling for family and individual traits, as well as for characteristics of the schools students attended between the daycare age and the age they took the Math/ language exams is crucial for correctly indentifying the causal effect of DC attendance on test scores. In the rest of this section we will show evidence that daycare attendance *is* correlated with other potential determinants of test scores, and mention some other problems that could influence our conclusions.

2.1 Family and the household

Table 2 contains information about the importance of parental input to the child's skill formation, as well as the relation of these inputs to daycare attendance. Besides the proportion of DC attendees displayed in the first column, this table also shows the proportion of children who had access to some Early Childhood School (ECS), being it either a daycare center or kindergarten, the average test score of former DC attendees and non attendees, and the sample distribution of parental inputs. As we see, children of more educated parents are more likely of having attended daycare centers and EC schools. The difference between test scores of DC attendees and non-attendees seems to depend crucially on the educational level of the parents, with the (absolute) magnitude being smaller than 0,1 sd for all of the families with parents below college level (in the lower tail it is even negative). Regarding to the presence of the parents living at home, we see that the probability the children attended DC in the early childhood does not depend on this factor, but the difference in test scores between DC attendees and non-attendees is almost null for students with parents out of home, and around 0.15 sd for those with parents at home. Turning now to the parental attitude towards learning, we see that, on one hand, pro-active parents are only slightly more willing to send their children to daycare centers, whereas the DC differential in Math test scores are higher among children of parents who *not* attend school meetings, incentivate children to do homework or read, or advise children not to miss classes. This result is not completely surprising, as such attitudes may be the *result* (and not the *source*) of the child's performance at school (it is plausible that it is the parents of children with bad performance that are more likely to advise children not to miss classes, to read books, or to do homework).

The next table presents differences in access to daycare centers across groups with different levels of wealth as measured by the size and number of members of the household, and by the availability of durable goods in the household. The results suggest that children living in more spacious and less populated households have higher probability of having attended daycare centers and kindergarten, and that the DC differential is greater in such households. Similarly, the differ-

ence in test scores between former DC attendees and non-attendees is negligible in households without some typical durable goods, but tend to be positive in wealthier households.

Overall, there seems to be some heterogeneity in the impact of daycare attendance on test scores across families with different parental input, being it especially strong among children of more educated and present parents.

2.2 School

As argued above, families more willing to invest their resources in their children's education may be both more likely to send them to DC and kindergarten, and to enroll them in better elementary schools. In principle, the quality of each school attended by the child should matter and be controlled for when measuring the net impact of DC attendance on test scores. Since we only have information about characteristics of the currently attended school, we can only investigate somehow the extent to which part of the DC differentials displayed in Table 1 may be due in fact to differences in schools children attended between DC and the moment they took the SAEB exam.

In Table 4a, we present how heterogeneity in pedagogical plan and infrastructure may be correlated with daycare attendance and influence the possible impact of this attendance on test scores. Regarding to the pedagogical plan, we notice that the probability that a student had a DC experience is similar in schools with different pedagogical policies, such as extra classes for delayed students, use of homework, or a specific program against repetition. Interestingly, schools that do not use textbooks have a higher proportion of DC attendees, at the same time that display a DC differential in test scores about 4 times bigger than that found in schools that do not use textbooks. In general, schools with no program against repetition and no program against dropout have higher DC differentials, but it may be the case that schools create such programs in response of bad performance of their students.

With respect to the infrastructure and ownership of the school, it is interesting to see that schools with a number of computers per student close to 1 are those with the highest proportion of former DC attendees, but are also those where being a DC attendee does not represent a difference in terms of test scores. In the same spirit, children currently enrolled in private schools, known as being far better than public schools in Brazil, are almost 3 times more likely to have attended daycare institutions, but neither within public schools nor within private schools DC attendance seem to produce an advantage in cognitive terms. These results suggest that an important part of the gross differentials observed between DC attendees and non-attendees is in reality due to the fact that DC attendees will more likely attend good elementary schools in the future whereas non-DC attendees will not.

Turning to the heterogeneity in the human capital of the school, we see that the principal's experience does not seem to be important to understand the effect of DC attendance on cognitive formation (in 98% of the schools the principal has at least 6 years of experience, and both DC attendance and the

DC differential are close of the overall pattern in these schools). Similar results hold with respect to the principal's education, but the conclusion is different when we focus on the period the principal has worked in the current school. In this dimension, higher tenure is often associated to a greater proportion of DC attendees and to a higher DC differential in test scores.

Regarding to characteristics of the teacher, we see that neither the experience of the teacher in the profession nor the time the teacher has tought that specific class, nor even the educational level, are related to the proportion of students who reported to have attended daycare centers in the class. The evidence slightly suggests that the DC differential may be bigger in classes taught by more qualified professionals (higher education and experience), but this pattern is not out of controversy.

2.3 OLS results

In this section, we will present OLS results, that can either be seen as an estimation of the possible causal effect of daycare attendance on test scores, in the case the decision of sending children to daycare centers is exogenous conditional on observables, or as one more set of descriptive statistics that aim to show what set(s) of determinants of test scores is more closely related to daycare attendance and could therefore confound the effect of daycare attendance on scores. The fact that one set of observable characteristics (family, school, etc.) is more related to daycare atendance may also suggest that we should be especially careful with potential unobservables that belong to that set. For this exercise, we estimated the equation:

$$y_{igt} = X_i\beta_{1t} + F_i\beta_{2t} + S_i\beta_{3t} + \delta_t DC_i + \varepsilon_{it}$$

where X_i denotes individual characteristics of the individual i; F_i and S_i refers to family and school characteristics, respectively; and DC_i si an indicator of previous attendance to daycare centers. Subscript g stands for the type of test taken (Math or language), and t is the grade (4th, 8th or 11th).

As displayed in Table 5, a direct comparison between Math test scores of former DC attendees and non-attendees shows that attendance to daycare is associated to 17 extra points in the final mark, or 0,35 standard deviations. After controlling for individual traits, this differential decreases to 14 points, mainly due to the inclusion of age in the regression[18]. A significant reduction in the coefficient associated to DC attendance occurs when we control for the family structure, especially maternal education, suggesting that part of the gross difference between scores of DC attendees and non-attendees is in fact due to the quality of parents as an input to child development. As discussed above, many explanations may help to understand this fact, such as strong preference for education among more educated parents, and higher opportunity cost of time of more educated parents. After controlling for maternal education and other parental attributes, the DC coefficient is cut by a half in the regression.

[18] See a discussion about the inclusion of age in the regression in the next subsection.

The DC coefficient remains stable as we include controls for parental attitude towards learning, wealth, an indicator of grade repetition, and the pedagogical plan of the school. It drops to about 1/3 of the original 17 when we include dummy variables characterizing the way the principal is chosen. Interestingly, the coefficient jumps up again when we include the principal's and teacher's traits. A possible explanation is that the best teachers are assigned to the worst behaved students (this fact is indeed reported in qualitative questionnaires aplpied to principals and teachers). If the main efefct of daycare attendance were to improve sociability and discipline, it is possible that former DC attendees end up having worse teachers, which ceteris paribus reduces their grades.

Finally, we see that the original differential shrinks to less than 1/4 if we include school fixed effects, becoming still signifficantly different from zero, but with magnitude smaller than 0.08 of a standard deviation. Most of this pattern is also observed if language scores are used instead of Math, with the main difference being that now school infrastructure seems to be closely related to daycare attendance.

One of the main difficulties in isolating the impact of daycare attendance on test scores is that most of the children who were sent to DC institutions also stayed in school for the kindergarten. One way of attempting to separate the effect of DC attendance from kindergarten attendance is to compare individuals with similar characteristics except that some started school in daycare centers whereas others started in kindergarten. We do this by including indicators of kindergarten attendance in the OLS regressions, therefore differentiating non-DC attendees in two groups (with and without kindergarten). In this setup, we see the DC variable as an indicator of having received both DC and kindergarten treatments, while the kindergarten dummy would discriminate those who have only received kindergarten[19].

The first thing to notice is that the gross DC differential declines to about a half if we include a kindergarten indicator (0,17 sd). This coefficient again drops to a half after the addition of family attributes. Similar to the pattern observed in the previous paragraph, most of the school characteristics not related to the principal's and teacher's traits make the coefficient to decerase further, whereas such traits help the coefficient to go up. After the inclusion of school fixed efefcts, the coefficient becomes no longer statistically significant.

To summarize this discussion, we should say that most of the DC gross differential observed in the data can be attributed to strong correlations between this indicator and other family and school inputs that directly affect test scores. In particular, families with highly educated parents are more willing to send children to daycare centers, and families that had sent children to DC also tend to choose better elementary schools for their kids as well as to use kindergarten

[19] Actually, children is asked to answer what their first school experience was. In principle, it is possible that some of the individuals who answered "daycare" stayed out of school at the kindergarten age. The fact that attendance rates in kindergarten are tipically 2,5 times bigger than in daycare centers, together with the evidence from the 1996/97 Living Standards Survey (PPV, IBGE. See Barros et all ???) that most of the former DC attendees were also kindergarten attendees make this possibility likely to be rare in our sample.

services. After controlling for all of these confounding variables and multiple treatments, it seems that the net effect of daycare attendance is either non significant or at least very small in magnitude.

2.4 Endogeneity of age

The inclusion of age as a regressor in an exercise aiming to estimate the causal effect of daycare attendance on test scores is not free of controversy. Although we do include age in this paper (mostly because we find its exclusion may lead to even more serious problems), we intend to briefly discuss some of the problems we devise and which have been neglected in most of the related literature.

If we admit that one of the possible effects daycare attendance may have on learning operates through making children be at the correct grade for her age (either because of non-cognitive channels, as motivation and endurance, or because of cognitive channels, since good marks are usually a pre-requisite for grade promotion), then once we control for age we are comparing two children that achieved the same grade at the same age with a difference that one of them had access to daycare, which in principle would give an advantage to the former DC attendee. In this case, either the non-DC attendee were unobservably outstanding so that she could compensate the lack of DC experience in another way in order to be in the same grade at the same age, or were the DC attendee unobservably not so good so that she achieved that grade at the same time than the non-attendee in spite of her original advantage. In other words, age in this case may be an endogenous variable in our sample of 4th graders.

The Graph 1 below helps us to see this fact by comparing the age distribution of former DC attendees and non-attendees in different grades. In the Brazilian system, the correct age for being at the 4th, 8th and 11th grades are 10, 14 and 17 years old, respectively. It is clear from this figure that children who had access to some type of early childhood education (daycare and/ or kindergarten) are significantly more likely to be at the correct age for the grade. While 63% of former DC/ kindergarten attendees had at most 10 years old at the 4th grade, only 40% of the children with no preschool experience were in this situation. Things become even more dramatic in the 8th grade, when the difference is 48% vs 28% in favor of former DC/ kindergarten attendees.

Again, there is a chance that the gross effect found in the graph above is due to either other (individual/ family) characteristics former DC/ kindergarten students have, instead of the attendance itself, or to other treatments that DC/ kindergarten attendees are more likely to receive, such as better elementary schools. In order to disentangle these confounding factors, we run a series of probit regressions with an indicator of being at the correct age for the 4th grade as the dependent variable (in these regressions, we also included 11 year-old children in the correct age group, but the results do not change if we restrict it to 10).

The middle column is the one of most concern to us, which refers to the partial effect of daycare attendance in a set of regressions that include kindergarten indicators as well. The first fact worth to mention is that even the gross

partial effect found in a regression with no controls is relatively small (5% out of 58% of children at the correct age). This number falls to less than 3% after the incluson of family and individual attributes, and becomes statistically insignificant once we control for school characteristics. This result is important as it suggests we are probably not incurring in big mistakes by including age in our exercises as a regressor. Treating age as an endogenous variable would require a careful search for instruments that are not obvious, and excluding age from the regressions would lead to biased conclusions, as it seems to be both an important determinant of test scores and correlated with daycare attendance.

Although we proceed treating age as an exogenous regressor, we do think more research should be done to verify the role this variable plays in the empirical strategy of articles that estimate effects of ECD interventions on later outcomes. We should always have in mind that one of the events that our exercises are mandatorily conditioned on is attendance to a given grade of the regular educational cycle, and not all of the children effectively achieve such grade or do this at the correct age.

3 Endogenous choice of school

The primary goal of our analysis is to estimate what would be the causal impact of daycare and preschool attendance on human capital formation, as (noisily) measured by standardized test scores applied at different points of the educational cycle. The main difficulty in doing so is to control for the fact that individuals who attended these school levels may be unobservationally different from those who started studying only at primary school. In this section we build a model that describes the process of human capital formation during childhood, in which the parental decision of sending children to school is endogeneous as well as the quality of the school attended.

Formally, we believe that test scores measured at grade g, age t are related to the level of human capital of individual i, υ_{igt},through:

$$\begin{aligned} y_{igt} &= G_g\left(\ln \upsilon_{it}\right) + u_{igt} \\ & \upsilon_{it} \perp\!\!\!\perp u^m_{igt} \end{aligned}$$

It is reasonable to assume that the current ability of an individual depend on family and school investments (proxied by their respective characteristics), and on her own past ability:

$$\upsilon_{it} = \upsilon^{\rho}_{it-1} H\left(F_{it}, S_{it}\right)$$

To illustrate the endogeneity problem that may arise in the estimation of the causal impact of daycare and preschool attendance on human capital, let us consider the linear-in-parameters form:

$$y_{igt} = \delta_0 + \delta_1' g_1\left(F_{it}, S_{it}\right) + \delta_2 \ln \upsilon_{it-1} + u_{igt}$$

Now, if each period at home increases the child's human capital at rate λ_H, whereas a period at school increases her human capital at rate λ_S, we have that a child that entered school at age a^* should have a level of human capital:

$$\ln \upsilon_{it-1} = \ln(\upsilon_{i0}) + (\ln \lambda_S)(t - a_i^*) + (\ln \lambda_H) a_i^*$$

and in this case the causal effect of starting school at age a^* (instead of $a^* + 1$) on $\ln \upsilon_{it-1}$ would be $\ln(\lambda_S/\lambda_H)$. This structure would deliver a reduced form relating test scores and the parameters of interest of the kind:

$$\begin{aligned} y_{igt} &= \delta_0 + \delta_1' g_1(F_{it}, S_{it}) + \delta_2 (\ln \lambda_S) t + \delta_2 (\ln \lambda_H/\lambda_S) a_i^* + \varepsilon_{igt} \\ \varepsilon_{igt} &= \delta_2 \ln \upsilon_{i0} + u_{igt} \end{aligned}$$

The central problem in our analysis is that families decide the age they want to send children to school, which means that $a^* = a^*(\upsilon_{i0}, Z)$, where Z includes part of the family attributes F, as well as other variables such as the past educational inputs provided to the children, S, the opportunity cost of time of family members and the cost of acquiring educational inputs to their children. As a result, a direct estimation of $E[y_{igt}|F, S, t, a^*]$ would result in biased estimates of $\ln(\lambda_S/\lambda_H)$, precisely because even if we assume that $\mathrm{E}(\upsilon_{i0}|F) = 0$ (or the stronger assumption $\ln \upsilon_{i0}|F, S, a^* \sim N(0, \sigma^2)$), it would not follow that $E(\ln \upsilon_{i0}|F, S, t, a^*) = 0$.

3.1 The choice of school

In our model, families are the decision makers, which value both current consumption, c, and future human capital of their children, υ', through a utility function of the type:

$$U(c, \upsilon') - d\mathbb{C}(Z)$$

where $\mathbb{C}(Z)$ denotes the cost, in terms of utility, of sending children to school (this includes the risk of getting a disease, being hurt, the psychic cost of letting a strange to take care of your children, etc, which may depend on a general set of variables Z).

There are two technologies of human capital accumulation, one that uses only household inputs, and the other which includes school inputs as well:

$$\begin{aligned} \upsilon' &= H_1(\upsilon, \tau F; \psi_1) \\ \upsilon' &= H_2(\upsilon, \tau F, S; \psi_2) \end{aligned}$$

where υ is the current level of human capital of the children, F denotes the productivity of family inputs per unity of time, τ is the time spent by the parents with their children, and S is an index of school inputs (φ_1, φ_2 are parameters that characterize these functions, which may vary with the child's age).

Families' decisions are also constrained by their budgets, $pc+qS=w(1-\tau)$, in which $S=0$ occurs if the family opts by not to send the child to school.

The maximization problem of the family can then be written as:

$$\begin{aligned} & \max_{d,c,S,\tau} U(c,\upsilon') - d\mathbb{C}(Z) \\ s.t \quad : \quad & pc+qS = w(1-\tau) \\ : \quad & \upsilon' = dH_2(\upsilon,\tau F,S;\psi_2) + (1-d) H_1(\upsilon,\tau F;\psi_1) \end{aligned}$$

and the general solution predicts that, if the child goes to school:

$$\frac{\partial H_2(\upsilon,\tau F,S;\psi_2)}{\partial(\tau F)} F = \frac{w}{q} \frac{\partial H_2(\upsilon,\tau F,S;\psi_2)}{\partial S}$$

which does not depend on the functional form of U (meaning that the optimal usage of inputs once the child goes to school depends only on the relative price of the inputs - time and school quality - and upon the technology, H_2).

To make the model simple, we assume logarithmic preferences and linear cost

$$U(c,\upsilon') = \ln c + \varphi \ln \upsilon' - d\theta Z$$

and a Cobb-Douglas technology,

$$\begin{aligned} H_1(\upsilon,\tau F;\psi_1) &= \Psi \upsilon^{\rho_1} (\tau F)^{\beta_1} \\ H_2(\upsilon,\tau F,S;\psi_2) &= \upsilon^{\rho_2} S^{\alpha} (\tau F)^{\beta_2} \end{aligned}$$

to get:

$$\begin{aligned} c_2^* &= \frac{1}{1+\varphi(\alpha+\beta_2)} \frac{w}{p} \\ \tau_2^* &= \frac{\beta_2 \varphi}{1+\varphi(\alpha+\beta_2)} \\ S_2^* &= \frac{\varphi\alpha}{1+\varphi(\alpha+\beta_2)} \frac{w}{q} \end{aligned}$$

; if the child goes to school, and:

$$\begin{aligned} c_1^* &= \frac{1}{1+\varphi\beta_1} \frac{w}{p} \\ \tau_1^* &= \frac{\varphi\beta_1}{1+\varphi\beta_1} \end{aligned}$$

if she does not.

The decision about whether to send children to school or not depends on the comparison between the maximum payoff obtained in each of these situations. If the family opts not to use school, then the utility associated to this choice is:

$$U^1(w,q,p,\upsilon,F) = \varphi A_1 + \ln\left(\frac{w}{p}\right) + \rho_1 \varphi \ln \upsilon + \beta_1 \varphi \ln F$$

If the child goes to school, the associated payoff is[20]:

$$U^2(w,q,p,\upsilon,F) = \varphi A_2 + \ln\left(\frac{w}{p}\right) + \varphi \rho_2 \ln \upsilon + \varphi \alpha S^* + \varphi \beta_2 \ln F - \theta Z$$

which means that the child is sent to school iff:

$$(A_2 - A_1) + \alpha \ln S^* + (\rho_2 - \rho_1) \ln \upsilon + (\beta_2 - \beta_1) \ln F > \theta Z$$

3.2 A 3-period problem

If we extend the previous setup to a multiperiod situation, we will end up with a simple dynamic model where individuals are sent to school at age a^* if

$$\Delta A_{(a^*-1)} + \alpha_{a^*-1} \ln S^*_{a^*-1} + \Delta \rho_{(a^*-1)} \ln \upsilon_{a^*-1} + \Delta \beta_{(a^*-1)} \ln F_{a^*-1} < \theta Z$$
$$and$$
$$\Delta A_{a^*} + \alpha_{a^*} \ln S^*_{a^*} + \Delta \rho_{a^*} \ln \upsilon_{a^*} + \Delta \beta_{a^*} \ln F_{a^*} > \theta Z$$

Assuming the technology and preferences are the same throughout the life-cycle, as well as the family characteristics, entrance in school at age a^*:

$$\Delta A + \alpha \ln S^* + \Delta \rho \ln \upsilon_{a^*-1} + \Delta \beta \ln F < \theta Z$$
$$and$$
$$\Delta A + \alpha \ln S^* + \Delta \rho \ln \upsilon_{a^*} + \Delta \beta \ln F > \theta Z$$

We will now consider the case in which the periods of the model correspond to the educational cycles, being the first period the interval between birth and 3 years-old (for which the school decision correspond to what we will generically call "daycare" attendance), the second period associated to the kindergarten phase (4-6 years old), and the third period being the first part of elementary school (1^{st} to 4^{th} grades, or 7-10 years-old[21]).

[20] $A_1 = \frac{1}{\varphi} \ln \left[\Psi^{\varphi} \left(\frac{\varphi \beta_1}{1+\varphi\beta_1} \right)^{\varphi \beta_1} \frac{1}{1+\varphi\beta_1} \right]$

$A_2 = \frac{1}{\varphi} \ln \left(\frac{\varphi^{\frac{\varphi(\alpha+\beta_2)}{1+\varphi(\alpha+\beta_2)}} \alpha^{\frac{\varphi\alpha}{1+\varphi(\alpha+\beta_2)}} \beta_2^{\frac{\varphi\beta_2}{1+\varphi(\alpha+\beta_2)}}}{1+\varphi(\alpha+\beta_2)} \right)^{1+\varphi(\alpha+\beta_2)}$

[21] The Brazilian educational system has changed in 2007, with the minimum age from which children must be in school had been changed from 7 to 6 years old. We will work as if the mandatory minimum age were still 7 years, because our data correspond to a cohort that were already in the 4^{th} grade in 2007).

In this case, children with [22]

$$\begin{aligned} \ln \upsilon_3 &= a_0^1 + a_1^1 \ln F + \alpha \ln S_3 + a_2^1 \ln \upsilon_0 \\ \ln \upsilon_0 &\geq \pi_{00} + \pi_{01} \ln F - \theta Z \end{aligned}$$

are those who attend daycare institutions. Children who start school at kindergarten must satisfy[23]:

$$\begin{gathered} \ln \upsilon_3 = a_0^2 + a_1^2 \ln F + \alpha \ln S_3 + a_2^2 \ln \upsilon_0 \\ \pi_{10} + \pi_{11} \ln F - \theta Z < \ln \upsilon_0 \leq \pi_{00} + \pi_{01} \ln F - \theta Z \end{gathered}$$

and finally, those who do not attend any type of ECD educational institution satisfy[24]:

$$\begin{aligned} \ln \upsilon_3 &= a_0^3 + a_1^3 \ln F + \alpha \ln S_3 + a_2^3 \ln \upsilon_0 \\ \ln \upsilon_0 &< \pi_{10} + \pi_{11} \ln F - \theta Z \end{aligned}$$

3.3 An estimable structural model

The 3-period model outlined above suggests that the decision of sending children to school in any period depends on a linear index of family characteristics (F), and on a set of variables Z that not necessarily have a direct effect on human capital accumulation.

In our data, we observe characteristics of the family and of the current school the student is enrolled (4^{th} graders), together with the age the student was first sent to school and measurements of abilities in Math and Language coming from

---

22

- $\pi_{00=} - \frac{\alpha}{\Delta\rho} \ln\left(\frac{w_0}{q_0}\right) - \frac{\Delta A}{\Delta \rho}; \pi_{01} = -\frac{\Delta\beta}{\Delta\rho};$
- $a_0^1 = \left[\beta_2 \ln \tau_2^* + \alpha\left(1 + \ln\left(\frac{\varphi\alpha}{1+\varphi(\alpha+\beta_2)}\right)\right)\right] \sum_{r=0}^{2} \rho_2^r;$

$a_1^1 = \beta_2 \sum_{r=0}^{2} \rho_2^r; a_2^1 = \rho_2^3$

23

- $\pi_{10} = -\frac{\alpha}{\rho_1 \Delta\rho} \ln\left(\frac{w_1}{q_1}\right) - \frac{1}{\rho_1\Delta\rho}\left(\Delta A + \Delta\rho \ln\left[\Psi\left(\frac{\varphi\beta_1}{1+\varphi\beta_1}\right)^{\beta_1}\right]\right)$

$\pi_{10} = -\frac{1}{\rho_1\Delta\rho}\left(\beta_1 \Delta\rho + \Delta\beta\right).$

- $a_0^2 = \alpha\rho_2 \ln\left(\frac{\varphi\alpha}{1+\varphi(\alpha+\beta_2)} \frac{w_2}{q_2}\right) + \rho_2^2 \ln \Psi + \beta_2\left(1+\rho_2\right) \ln \tau_2^* + \rho_2^2 \beta_1 \ln \tau_1^*$

$a_1^2 = \beta_2\left(1+\rho_2\right) + \rho_2^2\beta_1; a_2^2 = \rho_2^2 \rho_1$

24

- $a_0^3 = \left(\ln \Psi + \beta_1 \tau_1^*\right) \sum_{r=0}^2 \rho_1^r$
- $a_1^3 = \beta_1 \sum_{r=0}^2 \rho_1^r; a_2^3 = \rho_1^3$

test scores. We do not directly observe neither the opportunity cost of time , w, nor the cost of a quality adjusted unit of the educational input, q, but if we did, these could be good proxies for the quality of the school attended before the time the test is taken (past educational inputs are not observed either). We have two sets of instruments that could be used to proxy for the utility cost of sending children to school: measures of supply of daycare and kindergarten services per municipality (both public and private), and the incidence of contagious diseases in the population of 0-3 and 4-6 children. According to our model, these variables should shift the family choice for school without directly affect test scores. The simplest way to make our model coherent with the available data is to assume

$$\mathbb{C}(Z) = \theta' Z$$

where Z comprises school supply and contagious diseases variables per municipality. With these assumptions, we end up with the equations:

$$\begin{aligned} y_{g3} &= a_0^3 + a_1^3 \ln F + \alpha \ln S_3 + d_c \left[\left(a_0^1 - a_0^3\right) + \left(a_1^1 - a_1^3\right) \ln F \right] + \\ & \quad d_k \left[\left(a_0^2 - a_0^3\right) + \left(a2 - a_1^3\right) \ln F \right] + \varepsilon_{g3} \\ E\left[d_c | X, Z\right] &= \Pr\left[\ln \upsilon_0 \geq \pi_{00} + \pi_{01} \ln F - \theta Z\right] \\ E\left[d_k | X, Z\right] &= \Pr\left[\pi_{10} + \pi_{11} \ln F - \theta Z < \ln \upsilon_0 \leq \pi_{00} + \pi_{01} \ln F - \theta Z\right] \\ \varepsilon_{g3} &= \left[\left(a_2^3 + d_c \left(a_2^1 - a_2^3\right) + d_k \left(a_2^2 - a_2^3\right)\right) \ln \upsilon_0 + u_{g3}^m \right] \end{aligned}$$

4 Empirical Strategy

The theoretical model presented in the last section will provide guidance to correct the endogeneity problem that may arise with the inclusion of daycare attendance in test score regressions. Using the proper definition of the parameters, and matching the variables of the model with the data, such that $F_i = x_{igs}$ is a vector of student's attributes which includes his/ her family background; and $S = [x_{gs}, x_s]$, where x_{gs} is a vector of classroom's characteristics and x_s is a vector of school's characteristics, the last equation of our theoretical model stated that:

$$y_{igs} = \beta x_{igs} + \eta_1 dc_{igs} + \eta_2 dk_i + \gamma s_{gs} + \varepsilon_{igs}$$

where y_{igs} denotes test scores in math of child i in classroom g in school s, dc_i is a dummy variable that equals to 1 if the child started school at daycare and 0 otherwise, dk_i is a dummy variable that equals 1 if the child is at school at kindergarten and 0 otherwise, and ε_{igs} are the unobservable factors.

As pointed out before, ε_{igs} includes the initial ability of child i that is correlated with the parents decision to enroll his/her child in daycare. To deal with this endogeneity problem, we will use a control function approach proposed by Newey. Powell and Vella (1999). Based on the our theoretical model, the parental decision is related to the supply and quality of daycare centers they

face, and children's and family's traits. The parents decide to enroll their child at the first time at daycare, in kindergarten or after that if

$$\begin{aligned} dc_i &= \mathbf{1}\{\upsilon_{igs} > \theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}\} \\ dk_i &= \mathbf{I}\{\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs} \leq \upsilon_{igs} \leq \theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}\} \\ de_i &= \mathbf{I}\{\upsilon_{igs} < \theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs}\} = 1 - dc_i - dk_i \end{aligned}$$

where zc_i is a vector that includes variables related to supply and demand for daycare, zk_i is a vector that includes the same variables related to the supply and demand for kindergarten, υ_{igs} is the unobservable factor. We assume that the variables included in zc_i and zk_i do not have a direct impact on children's outcome at the 4^{th} grade of elementary school.

We understand this problem as a selection problem framework. Controlling for family characteristics, the child's initial ability and parents' needs should drive the parental decision about the age children should start at school. Parents' needs may include whether they need to work outside, parents' preferences for education, etc. At the end, some children will be selected to start school before kindergarten, and others will enter school later. The vector z works as a vector of instrumental variables that allows us to isolate the exogenous part of this decision, and to use a function of the endogenous part to control for selection.

To deal with this selection problem, we impose some assumptions about the error terms. The first one assumes that errors are independent of $z \equiv (zc, zk)$ and that all the selection problem is related to the decision of enrolling i in daycare or not

A1:

$$(\varepsilon_{igs}, \upsilon_{igs}) \perp (zc_i, zk_i, x_{igs}, s_{gs})$$

Under assumptions A1, we have

$$\begin{aligned} & \mathbb{E}\left[y_{igs} | x_{igs}, s_{gs}, dc_i, dk_i, z_i\right] \\ = \; & \beta x_{igs} + \eta_c dc_i + + \eta_k dk + \gamma s_{gs} + \mathbb{E}\left[\varepsilon_{igs} | x_{igs}, dc_i, dk_i, z_i\right] \end{aligned}$$

Let us assume that in period 1 the parents decide to enroll the kid's in daycare or not. If they enroll their kid at daycare, the child stays at school at kintergarten. If they decide not to enroll their child at daycare, then they solve another maximization problem in the second period. Notice that we have 3 possible situations: the child started school at daycare, $dc_i = 1$ and stays at school; the child started school at kindergarten, $dk_i = 1$ and the child started school after kindergarten, $dk_i = 0$.

$$\begin{aligned} & \mathbb{E}\left[\varepsilon_{igs} | x_{igs}, dc_i, dk_i, s_{gs}, z_i\right] \\ = \; & \mathbb{E}\left[\varepsilon_{igs} | dc_i = 1, x_{igs}, z_i, s_{gs}\right] \cdot dc_i \\ & + \mathbb{E}\left[\varepsilon_{igs} | dk_i = 1, x_{igs}, z_i, s_{gs}\right] \cdot dk_i \\ & + \mathbb{E}\left[\varepsilon_{igs} | de_i = 1, x_{igs}, z_i, s_{gs}\right] \cdot (1 - dc_i - dk_i) \end{aligned}$$

To find a function that controls for selection, we also impose that:
A2:

$$\begin{aligned} (\varepsilon_{igs}, \upsilon_{igs}) &\sim N(0, \Omega) \\ \mathbb{V}ar[\upsilon_{igs}] &= 1 \\ \mathbb{C}ov[\varepsilon_{igs}, \upsilon_{igs}] &= \sigma_{v\varepsilon} \end{aligned}$$

Under assumptions A1 and A2,

$$\begin{aligned} \mathbb{E}[\varepsilon_{igs}| dc_i = 1, x_{igs}, z_i, s_{gs}] &= \mathbb{E}[\varepsilon_{igs}| \upsilon_{igs} \geq \theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}, x_{igs}, z_i, s_{gs}] \\ &= \sigma_{v\varepsilon} \cdot \frac{\phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs})}{1 - \Phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs})} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}[\varepsilon_{igs}| dk_i = 1, x_{igs}, z_i] &= \mathbb{E}[\varepsilon_{igs}| \theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs} \leq \upsilon_{igs} \leq \theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}, x_{igs}, z_i] \\ &= \sigma_{v\varepsilon} \cdot \frac{\phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}) - \phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})}{\Phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}) - \Phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\varepsilon_{igs}| dk_i = 0, x_{igs}, z_i] &= \mathbb{E}[\varepsilon_{igs}| \upsilon_{igs} \leq \theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs}, x_{igs}, z_i] \\ &= -\sigma_{v\varepsilon} \cdot \frac{\phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})}{\Phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})} \end{aligned}$$

At the end, we have that the linear-in-parameters selection-corrected model is

$$\begin{gathered} \mathbb{E}[y_{igs}| x_{igs}, dc_i, x_{gs}, x_s, z_i] \\ = \beta x_{igs} + \eta dc_i + \alpha x_{igs} \cdot dc_i + \gamma x_{gs} + \zeta x_s \\ + \sigma_{v\varepsilon} \Im(zc_i, zk_i, x_{igs}, s_{gs}) \end{gathered}$$

where the control function is built such that:

$$\begin{aligned} \Im(zc_i, zk_i, x_{igs}, s_{gs}) = \sigma_{v\varepsilon} \Bigg[& \frac{\phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs})}{1 - \Phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs})} \cdot dc_i \\ & \frac{\phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}) - \phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})}{\Phi(\theta_c zc_i + \rho_c x_{igs} + \alpha_c s_{gs}) - \Phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})} \cdot dk_i \\ & - \frac{\phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})}{\Phi(\theta_k zk_i + \rho_k x_{igs} + \alpha_k s_{gs})} (1 - dc_i - dk_i) \Bigg] \end{aligned}$$

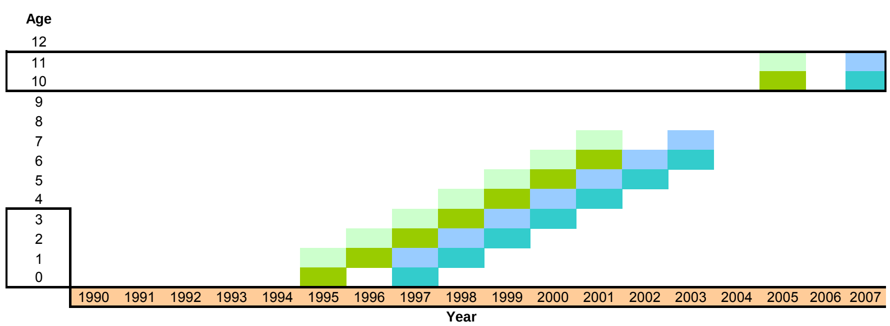
The model is estimated by Maximum Likelihood.

5 Results

In this section, we present our results. The descriptive analysis shows that attendance to daycare may be related to parents' preferences for education and with

the quality of daycare and kintergarten instituions. In this section, we deal with this potential endogeneity by using the control function approach presented in the methodology section. As instruments (exclusion restrictions in the daycare and kindergarten participation equations), we use the number of unused classrooms in urban schools with daycare indicators (proxying the exogenous portion of daycare supply) along with the percentage of kids between 0 to 3 years old hospilized because of respiratory infection diseases (IRA) and by intestinal diseases, and the he percentage of kids between 4 to 6 years old hospilized because of respiratory infection diseases (IRA) and by intestinal diseases (which should in principle affect directly the family's decision). Both instruments vary at the municipality level[25].

The students that were at the the last of elementary school in 2005, we could have started daycare 10, 9, 8 or 7 years ago. We use the instruments at the municipality level since 1995. The Lexis diagramas illustrates which time period of the instruments we are considering for the two cohorts of students: with 10 or 11 years old at 2005 and with 10 or 11 years old at 2007.



The exclusion restriction is that avalability of classrooms and the spread of infectious diseases only affects test scores of the students at last year elementary school by affecting the parent's decision to enroll their kids for the first time at school at daycare and kintergarten.

5.1 Estimations

At the individual level, we use SAEB 2005 and estimate the models presented in section **??**. Table 8 presents the results of the first stage when we relate the probability of starting school at daycare or kindergarten with individual's characteristics, principal and teacher's characteristics, school's characteristics and the vector of instruments. The first column contains the coefficients of the control function, which represent the determinants of the family decision to enrool their children for the first time either at 0-3 years old (the proper age to be in a daycare institution in Brazil), or at 4-6 years old (the proper age

[25]The risk of getting contagious diseases at the school is one of the main concerns parents have to enroll their children in these institutions.

for kindergarten), whereas columns 2 do the same for an indicator of kindergarten attendance. In column 3 and 4, we repeat the exercise by modelling the probability of beginning school at daycare/ kindergarten in a linear way. The predicted values that emerge from models 3 to 4 will then replace the endogeneous regressors in **??** in a 2SLS estimation that aim to check the robustness of our control function strategy to assess the impact of ECD education on test scores.

Table 8 shows that parent's education and family socio-economic background[26] are important determinants of the probability of starting school at daycare or kindergarden, hence confirming the evidence displayed in the descriptive analysis. In addition, being enrolled in a private school at the 4^{th} grade has a positive impact on the probability of attending daycare. In our exercise, the school variables may be not only controlling for a direct have school quality has on test scores, but also serving as a proxy for the quality of previous schools children attended, as we do not observe these attributes. In fact, both the descriptive analysis and coefficients associated to the family choice for the age to start school show that families who enroll their children in better schools at the 4^{th} grade are also more likely to have used daycare services in the past (even though type of school has a negative impact on the probability of starting school at kindergarten. One possible explanation, it is that parent that have access to good quality daycare rather choose to send their kids to school earlier).

On Table 9, we present the estimations of the main equation of our model, relating daycare and kindergarten attendance with Math test scores. The first column comes from an OLS regression. If the daycare and kindergarten decisions were exogenous, the coefficients displayed in this column would provide the daycare and kindergarten effects we are looking for. The second column shows our favorite results based on the control function regressions described on the previous section, whereas the last column contains the linear IV regressions using the same set of instruments as the second one. On the bottom of this table we present a statistical test of the null hypothesis that starting school at daycare and kindergarten have the same impact on test scores. Since most of the children that start school at daycare stay at school for kindergarten, a non-rejection of the null hipothesys would mean that we have no evidence that daycare attendance has a positive impact net of kindergarten attendance.

As column OLS of Table 9 shows daycare and kindergarten have a positive efffect on math test scores, when we do not control for endogeneity, although apparently children that attended daycare centers perform almost equally well as children who entered school only at kindergarten. These effects do not disappear when we control for endogeneity, however the second column show we have evidence of selectivity in the decision of when to enroll the children at school. These results indicate that attending daycare and lindergarteb is proba-

[26] Since we do not have information about family income in the data, we follow the literature and create a index by factorial analysis that includes 12 itens of the questionnaire associated with family access to consumer goods, like tv, car and infrastruture of the household, like number of bathrooms, if they have a maid, etc.

bly related to parent's preferences for education and availability of good quality daycares and schools. The F-tests of the equality of the daycare and kindergarten coefficients shows that children who started school at daycare perform a little better than children who start at kindergarten. If both of them receive the kindergarten treatment (which is likely to occur as most children who start at daycare stay in school at kindergarten), we conclude that the net effect of daycare attendance is positive and small. It is interesting to notice that the magnitude of the kindergarten and daycare coefficients, around of 11, represents 0,25 of the standard deviation of Math test scores (see Table 1). This number is very similar to that found by Berlinsky et al (2009) for the impact of kindergarten attendance in Argentina (of 0,23 s.d.), and to that found by Felicio et al (2009) for a special sample of students of Sertaozinho (0.25 s.d.)[27]. As we will see in the next subsection, our estimations using a panel of schools with fixed effects for the schools deliver the same magnitudes for this effect. The regressions using standard IV estimators seem not very suited to our purposes, maybe due to the fact that the linear projection of the first stage incurs in the same type of problems that the predicted values of linear probability models do. The effect of kindergarten attendance becomes abnormally high (1,5 sd), and some coefficients contradict the evidence found elsewhere (e.g. the father's education becomes much more important than the mother's in explaining test scores). In this scenario, daycare attendance has a positive (significant) impact on test scores and the kindergarten effect is negative and significant.

5.2 Robustness using school-level data

Although we consider the structural approach used do far as the most promising use of the available data we have, people may argue that the conclusions we found are not robust to changes in the functional forms we assumed in the structure of the model, or that endogeneity may still be in place, if for example the daycare and kindergarten indicators were in fact proxies for the whole educational trajectory of the children (one could imagine that families who chose to send children earlier to school value education the most. If this is the case, these children may have received many other educational inputs before the 4^{th} grade that other children did not. The daycare and kindergarten indicators would in these cases capture more than the early childhood education effect, and endogeneity would arise if the valu placed by the family to education also affects test scores directly).

In order to address these issues, we pursue a completely different identification strategy to estimate the causal effect of early childhood education on test

[27]In this study, the authors applied a different exam to all students enrolled at the 2^{nd} grade of elementary school in Sertaozinho, a medium-size city in the countryside of the state of Sao Paulo. In this rich dataset, they collected information about the duration and exposition students had to daycare and kindergarten as well as a more detailed picture of the socioeconomic status of tehir families. Using a propensity score matching methodology, the authors found that attendance to kindergarten has an impact of 0,25 sd and that the duration does not matter (students that had only one year of preschool education performed equally well as those who had two, three and more than three years).

scores, yet controlling for part of endogeneity. In this second approach, we use a set of public schools that we can observe at two points in time. The advantage of this approach is that we can control for unobservable school characteristics that are constant over time and can be related to the amount of students in that school that went to daycare. As the theoretical model predicts, the available quality of education influences the decision of sending children to school at any point in time. As part of the quality of school is unobservable (e.g. the degree of parental involvement in administrative decisions), this may be an additional source of endogeneity in our problem, especially if (as we assumed before) children attend schools of a similar level of quality throughout the educational cycle (and in the school currently attended). Therefore, the inclusion of a school-level fixed-effect may help to mitigate this second source of endogeneity.

Inspired in **??**, we use a model very similar to the individual-level data model in which mean scores at the school level are a function of the average characteristics of the students in school, and of the characteristics of the school itself.:

$$\overline{y}_s = \beta\overline{x}_s + \eta_1\overline{dc}_s + \eta_2\overline{dk}_s + \gamma\overline{s}_{gs} + \overline{\varepsilon}_s$$

The proportion of students that attended daycare is a function of student´s parents decision, and is a function of the quality and supply of daycare in the neighborhood of that school and the characteristics of the students in school s.

$$\overline{dc}_s = \theta_c z_{cs} + \rho_c\overline{x}_{gs} + \gamma_c\overline{s}_s + \epsilon_{is}$$

Similarly, for kintergarten we have:

$$\overline{dk}_s = \theta_k z_{ks} + \rho_k\overline{x}_{gs} + +\gamma_k\overline{s}_s + \vartheta_{is}$$

As in the first strategy, we use a control function approach. The only difference is that we treat the endogenous variable in this case as a continuous variable (i.e. the proportions of children that started school at daycar and kindergerten, respectively). In this case,

$$\mathbb{E}\left[\overline{y}_s \middle| \overline{x}_s, \overline{dc}_s, \overline{dk}_s, \overline{x}_{gs}\right] = \beta\overline{x}_s + \eta_1\overline{dc}_s + \eta_2\overline{dk}_s + \gamma\overline{s}_{gs} + \mathbb{E}\left[\overline{\varepsilon}_s \middle| \overline{x}_s, \overline{dc}_s, \overline{dk}_s, \overline{s}_{gs}\right]$$

Under assumption A1

$$\begin{aligned}
\mathbb{E}\left[\overline{\varepsilon}_s \middle| \overline{x}_s, \overline{dc}_s, \overline{dk}_s, \overline{s}_{gs}\right] &= \mathbb{E}\left[\overline{\varepsilon}_s \middle| \overline{x}_s, z_{cs}, z_{ks}, \overline{s}_{gs}, \epsilon_s, \vartheta_s\right] \\
&= \mathbb{E}\left[\overline{\varepsilon}_s \middle| \epsilon_s, \vartheta_s\right] \\
&= \Gamma\left(\epsilon_s, \vartheta_s\right)
\end{aligned}$$

We impose a parametric functional form for $\Gamma\left(\epsilon_s, \vartheta_s\right)$

$$\Gamma\left(\epsilon_s, \vartheta_s\right) = \psi_1\epsilon_s + \psi_2\vartheta_s$$

We estimate this model using a two-step procedure. In the first step, we estimate the model for $\overline{dc}_s$ $\overline{dk}_s$ by ordinary least squares and obtain the residuals,

$\widehat{\epsilon}_s$ and $\widehat{\vartheta}_s$. In the second step, we estimate a regression that relates $\overline{y}_s$ with $\overline{x}_s$, $\overline{dc}_s$, $\overline{dk}_s$, $\overline{x}_{gs}$, $\overline{x}_s$, $\widehat{\epsilon}_s$ and $\widehat{\vartheta}_s$.

In a second exercise, we use a panel of public schools in Brazil and estimate the model presented in section **??**. The dependent variable is the difference in the average test scores in the 4^{ht} grade of these schools between 2007 and 2005, and we are interested in the effect of the difference in the percentage of kids that started school at daycare or kindergarten between these years. In this case, we may control part of the endogeneity of the daycare/ kindergarten decisions by just adding a school fixed-effect, or we can also built a control function at the school level if the inclusion of the fixed-effects does not suffice[28].

Table 10 shows the first stage of the panel-IV regressions, when we relate the percentage of kids that atttended daycare and kindergarten in each school with the characteristics of the students of the school, characteristics of the principal and of the teachers and the vector of instruments. This table shows age and mother's education are important determinants of the percentage of students that attended daycare. The instruments are not jointly signficant, and we may have a problem of weak instruments.

Table 11 shows the impacts with and without instruments. When we do not control for the potential endogeneity of the daycare and kindergarten indicators, the results are similar to those obtained at the individual level both in magnitude (between 10 and 15 grading points) and significance (1%). The exercise controlling for endogeneity shows no significance of the DC and kindergarten variables. As mentioned before, the control function in this case is just an aggregation of the linear IV model used in the last exercise of the previous subsection, and the problems discussed there may again be present now. Moreover, if the instruments are weak it may be the case that the pure fixed-effects estimation deliver less biased estimators than the IV strategy. The fact that the fixed-effects estimation almos coincide with the results obtained at the individual level (and match with other results found in the literature) indicate that they may be the right answer to our question.

6 Final Remarks

In this study we explored the data from SAEB, a Math and language exam applied to a sample of Brazilian students enrolled at the 4th, 8th and 11th grade of the basic educational cycle (the last grades of the elementary, junior high and high school, respectively), to investigate the impact of attendance to daycare and kindergarten institutions on the formation of cognitive skills. Our favorite results for Math exams of 4th graders show that children who had some kind of preschool experience perform around 0,25 standard deviation above those who did not attend either a daycare or a kindergarten school during early childhood.

[28] In such case, we need to use the difference in the instruments between the cohorts. For example, as an instrument of the demand of daycare for kids with 10 years old in 2005 and 2007, we use the difference in the ratio of hospitalization by IRA between 1997 and 1995. Notice that the inclusion of linear control functions will provide the standard linear IV estimators.

This result is similar in magnitude to those found by Berlinski at all (2009) in Argentinian 3rd graders (0,23 s.d.). It is interisting to notice, however, that we identify a small difference in performance between children who started school in a daycare center (up to 3 years old) and those who started at kindergarten (4-6 years old). Since most of the children who started school before kindergarten also attended kindergarten, it may be the case that the important event on a child's early childhood educational life is in fact the kindergarten attendance, instead of daycare itself (more research and better data should help to investigate this hypothesis). Our results also match the conclusions of Felício et al (2009), which used a more detailed dataset to estimate the impacts of daycare and preschool attendance on test scores for 2^{nd} graders among students of Sertaozinho (a medium-size city in the countryside of Brazil).

A major concern in our estimations was to deal with the potential problem of endogeneity in the family decision of sending children to daycare centers. It is plausible that such decision may depend on individual characteristics of the children (if for instance parents see DC institutions as a way of compensating delays in the child development), and family characteristics (on one hand, parents may have strong preferences for school, in which case children sent to daycare institutions may be also more likely to attend good kindergarten and elementary schools. On the other hand, parents with more human capital may transfer part of this capital to their children at the same time that are more likely to use DC institutions in order to have more available time to work). If children sent to daycare centers and kindergarten institutions differ sistematically from those that start school only at the elementary school in unobservable traits, we may attribute a wrong causal relationship to daycare and kindergarten attendance on test scores. We use a control function approach to correct for this potential endogeneity, using variables associated to the supply of daycare and kindergarten (which supposedly affect positively the family decision of using daycare/ kindergarten services without directly affect test scores in the future), and the incidence of contagious diseases among 0-6 years-old children (which we expect to affect the DC/ kindergarten decisions negatively), as exclusion restrictions. The results show that in fact children who started school at daycare and kindergarten are unobservably different than those who did not, in terms of test scores. Notwithstanding, the direct impact of daycare/ kindergarten on test scores changes very little once we controlled for the potential endogeneity.

In this study, two approaches that explore different types of variation in the data were used to estimate the causal effect of preschool attendance on Math test scores. The first approach used the variation in the probability that a child started school at daycare/ kindergarten induced by a variation in the exclusion restrictions associated with the supply of daycare/ kindergarten services at the municipality level as well as with the spread of contagious diseases among children at preschool age (also at the municipality level). The second approach explored the longitudinal variation in the proportion of the students that reported to have started school either in daycare or in kindergarten institutions at the school level (by introducing a school fixed-effect in a panel of observations at the school level). The fact that these two very different sources of variation

led to the very same magnitudes of the estimated impacts produce a convincing piece of evidence that this may be in fact the causal effect we were looking for.

7 Appendix A: Description of the data

In our analysis, we rely primary on two data sets: the Brazilian National Evaluation System of Basic Education (SAEB) and Prova Brasil. We use two auxiliary data sets to get our instrumental variables, the Censo Escolar (Educational Census) and DATASUS (Unified Health System database).

SAEB is a biennial survey conducted by the National Institute for Educational Studies and Research (INEP). SAEB evaluates students currently enrolled in the 4th and 8th grades of elementary school and in the 3rd year of high school in two subjects: Portuguese and Math. In this survey, students not only take standard exams in Portuguese and Math but also fill out a questionnaire that contains information about their socioeconomic status, behavior towards learning and parent's participation in the educational process. Teachers and Principals also answer contextual questionnaires on teaching practices, management and socioeconomic background. In addition, this survey collects information about the infrastructure in the school, for example, availability of textbooks to the students, if the classrooms have air circulation and enough light, etc.

SAEB collects information from a sample of students in urban and rural schools in all the states in Brazil. SAEB divides the population of schools into subpopulations based on location and size of the school (number of students in the last year of elementary school, middle school and high school). In each subpopulation, they randomly select schools to be part of the sample. In each school, they randomly select one or two classrooms of each grade to take the exams and fill the questionnaires. In each classroom, students are randomly selected to take the Portuguese or the Math exam. In our research, we will be especially concerned with the retrospective questions about daycare attendance, which were asked only to students in 2005 and 2007. Since SAEB is a sample of all schools in Brazil, it includes private and public schools.

Prova Brasil is a census of all public schools in Brazil. Prova Brazil uses SAEB´s proficiency tests and questionnaires. In 2007 and 2005, INEP conducted two surveys, one was a sample of schools (SAEB) and the other a census of public schools (Prova Brasil). Both surveys used the same questionnaires and tests, and their results are comparable. In this first draft, we use the information collected by SAEB 2005 and Prova Brasil 2007, and focus on tests scores of students in 4th grade who took the math exam. Using SAEB 2005 and Prova Brasil 2007, we could have a panel of public schools in these two years.

The SAEB 2005 sample has 41,783 students allocated in 3,607 classrooms in 2,811 schools. The math test scores will be our measure of achievement for students in the 4th grade of elementary school. We use sex, age, race, parent's education, the socioeconomic index and a measure of student's background in school as student's characteristics. To be precise, the vector of individual controls includes (i) a dummy variable that equals to 1 for females, (ii) age, a

discrete variable that varies from 8 until 15, where the last category represents students 15 years old or more; (iii) a dummy that equals to 1 if the student is white; (iv) an aggregate index that represents the socioeconomic level of the student's household; (v) parent's schooling, discretely measured according to the highest educational level achieved by each parent[29]. In addition, our main variable is a dummy that equals to 1 if the student attended daycare centers.

Since the data do not include an important characteristic of student's family background, the household income, we follow the literature and construct an aggregate index that represents the socioeconomic index of the students. This index is constructed by a Principal Component Analysis of 12 items in the questionnaire, which are related to the infrastructure and to the existence of some consumer durable goods in the household.

We consider the following school's and classroom's characteristics: (i) location (region of Brazil), (ii) type of school (dummy variable that equals to 1 for private, and 0 for public schools), (iii) characteristics of the teachers in each classroom (years of schooling, experience,as measured by the number of years as a teacher, race, and salary), (iv) total number of students in the classroom, (v) principal's characteristics (education, experience, as measured by the number of years as a principal in the school, and age), (vi) a variable that measures if the students in the classroom have books at home.

Table A1 shows the descriptive statistics of all the variables used in the exercises. It shows that 38% of our sample attended daycare, while 76% attended kindergarten. Only 37% of the students in the sample are white and 26% students have repeated a grade. 70% of schools are public and 36% are in the Northeast area. The mean classroom size is approximately 30 students, which can be considered big for international standards of elementary schools. Our instrumental variables vary only at the municipality level. In this first draft, we have only the school census information. Table 1 shows that the mean of enrollment per municipality is 200 students and there are some municipalities that have no daycare centers.

Graph A1 shows the distribution of tests scores. This distribution is slightly right-skewed, with mean of 189 points. The test was constructed in such way that the average student in the 4th grade should have a score of 250. Compared to this standard, the mean is very low and the majority of the students are below 250.

Graph A2 shows how test scores varies by previous attendance to daycare centers whereas Graph A3 displays the proportion of children that attended daycare centers by the educational level of the mothers. The first graph shows that children who attended daycare have higher test scores, while the second

[29] Possible educational levels are illiterated, less then elementary school, elementary school degree (4th grade), less than junior high school, junior high school degree (8th grade), less than high school, high school degree (11th grade), less than college, college degree, more than college degree. We used the middle point of each level to construct a discrete variable that could approximately measure the number of completed grades of each parent. Respondents that answered to have less than elementary school, for instance, were assigned the value of 2 years of schooling and so on.

shows that children of more educated mothers are more likely to have attended daycare. Potential explanations for these facts are that mother's education and time in daycare centers are complimentary inputs in the production of the child's human capital, or that the naive inclusion of daycare indicators in a regression would in fact capture the fact that more educated mothers are those who need to send their children to daycare centers, precisely because their opportunity cost of time is greater (in which case the daycare variable would be simply capturing part of the effect of maternal education the child ability). Hopefully the use of instruments will help us to disentangle these possibilities.

Graph A4 shows how test scores vary with the type of school. As we expected, students in private school have higher test scores than students in public schools. In Graph A5, we relate school type with daycare attendance. As we can see, almost 70% of the students that attended daycare went to private school. This graph shows that daycare can be associated with parents´ preferences for education. Parents that care about education enroll their kids early in school. Since we do not have a measure of family income, we use socioeconomic index as a proxy for it. Graphs A6 shows that the distribution of the socioeconomic index varies by attendance to daycare. The distribution for children that attended daycare has a higher median that the distribution for kids that have not attended daycare.

In the last graphs, we show the variability of our instruments. Graph A7 plots the average daycare attendance in each municipality against the average enrollment in each municipality 1997, and in Graph A8, we have the same plot for 1998[30]. These graphs show that the variability among the municipalities is not big, and we have a lot of municipalities that children went to daycare, but there were no public daycare centers. This results indicates that probably a part of our sample went to private daycare centers, and enrollment may not be a good instrument.

Our instruments for the supply side of daycare come from the Census of Basic Education (Censo Escolar) from 1997 until 2001. The Censo Escolar is a census of all public schools in Brazil up to high school. It also includes information about schools with special education and with a special program for illiterated adults (EJA). Of interest to us is the information about the size and the infrastructure of each school, for example, number of teachers and enrolled students by educational level, promotion and dropout rates, etc. The Censo Escolar is a key tool for the implementation of many public policies, such as food and transportation for students and the expansion of facilities in each school. In this paper, we will use the number of students enrolled in public daycare centers as proxies for the supply of this service. Since we focus on students at 4^{th} grade in 2005, we use the information from 1997 and 1998 censuses, collected when the 4^{th} graders were around 3 years old.

Finally, our instrument from the demand side comes from the Unified Health System (DATASUS). The DATASUS is the part of the health department re-

[30] In the y-axe of this graph, we have the percentage of students that have attended daycare in municipality s and the x-axe, we have the percentage of daycare enrollment in relation to total enrollment in municipality s.

sponsible for data collection about the heath system in Brazil, such as the number of individuals infected by different types of diseases, number of deaths per *causa-mortis*, and so on. In this study, we will use information about deaths provided by the Sistema de Informações sobre Mortalidade (SIM) and information about the incidence of various types of diseases as measured by the number of hospitalizations, provided by the Sistema de Informações Hospitalares (SIH-SUS).

The SIM information comes from the death certificates, which includes information about age, sex and cause of death, classified according to the CID (international diseases classification). From 1975 until 1995, SIM used CID-9, having switched to CID-10 in 1996. In this study, we will use information about the number of deaths of 0 to 3 years old children caused by respiratory infection diseases (IRA) or by intestinal diseases, aggregated at the municipal level.

SIH-SUS was created in 1990 and its main goals are to evaluate the heath policy in Brazil and to create a unified system with necessary information for the financial transfers to hospitals in the country. All of the public and private hospitals affiliated to the public heath system need to fill out a document for each hospitalization. With this document, this system can construct a data set that contains information about all the hospitalizations that are covered by the public heath system. This data has information about the diagnostic and medical procedures, demographic characteristics of the patients, infrastructure of the hospital, etc. In this paper, we are going to use the information about hospitalization of 0 to 3 years old children caused by respiratory infection diseases (IRA) or by intestinal diseases, per municipality.

8 References

References

[1] BARNETT, W.S. (1992). "Benefits of compensatory preschool education". Journal of Human Resources, v.27, N.2, 279-312.

[2] BEHRMAN, J.R., Y. Cheng, and P. Todd (2004). "Evaluating pre-school programs when length of exposure to the program varies: a nonparametric approach". Review of Economics and Statistics, v.86, N.1, 108–132.

[3] BERRUETA-CLEMENT, J., L.J.Schweinhart, W.S. Barnett, A.S. Epstein, and D.P. Weikart (1984). **Changed lives: The effects of the Perry Preschool Program on youths through age 19**. Ypsilanti, MI: The High/Scope Press.

[4] BLAU, D. and J. Currie (2006). "Pre-school, day care, and after-school care: who's minding the kids?". Handbook of the Economics of Education, v.2, 1163-1278. Elsevier, ed.

[5] BECKER, G. (1993). "Human capital: a theoretical and empirical analysis with special reference to education". 3^{rd} ed. The University of Chicago Press.

[6] BELFIELD, C., M. Nores, S. Barnett, and L. Schweinhart. (2006). "The High Scope/ Perry Preschool Program: cost-benefit analysis using data from the age-40 follow-up" Journal of Human Resources, v.41, N.1, 162-190.

[7] BERLINSKI, S., S. Galiani, and M. Manacorda (2008). "Giving children a better start: preschool attendance and school-age profile". Journal of Public Economics, v.92, 1416-1440.

[8] BERLINSKI, S., S. Galiani, and P. Gertler (2009): "The effect of pre-primary education on primary school performance," forthcoming, Journal of Public Economics.

[9] BORGHANS, L., A. Ducksworth, J. Heckman, and B. ter Weel (2008). "The economics and psychology of personality traits". NBER Working Paper N.13810.

[10] CAMPBELL, F., C. Ramey, E. Pungello, J. Sparling and S. Miller-Johnson (2002). "Early childhood education: young adult outcomes from the Abecedarian project". Applied Developmental Science, v.6, N.1, 42-57.

[11] CAMPBELL, F., E. Pungello, C. Ramey, S. Miller-Johnson and M. Burchinal (2002). "The development of cognitive and academic abilities: growth curves from an early childhood educational experiment". Developmental Psychology, v.37, N.2, 231-242.

[12] CAMPBELL, F., B. Wasik, E. Pungello, M. Burchinal, O. Barbarin, K. Kainz, J. Sparling, and C. Ramey (2008). "Young adult outcomes of the Abecedarian and 32 CARE early childhood educational interventions" Early Childhood Research Quarterly, v.23, 452-466.

[13] CARNEIRO, P. and J. Heckman (2003). “Human capital policy” In Heckman, J., A. Krueger, and B. Friedman Inequality *In*: **America: What Role for Human Capital Policies?**, MIT Press.

[14] CUNHA, F. and J. Heckman (2009). “The economics and psychology of inequality and human development”. NBER Working Paper no. 14695.

[15] CURRIE, J. and D. Thomas (1995) “Does Head Start make a difference?”. American Economic Review, v.85, N.3, 341-364.

[16] CURRIE, J. and D. Thomas (2000). “School quality and the longer-term Effects of Head Start”. Journal of Human Resources, v.35, N.4 755-774.

[17] FELICIO, F; R. Menezes and A.C. Zoghbi (2009). "The effects of early childhood education on literacy scores using data from a new Brazilian assessment tool". Working paper presented at the XXXVII National Meeting of Economics. ANPEC, Foz do Iguaçú (Brazil).

[18] MAGNUSSON, K., M. Meyers, C. Ruhm, and J. Waldfogel (2004). “Inequality in preschool education and school readiness. American Educational Research Journal, v.41, N.1, 115-157.

[19] MAGNUSSON, K., C. Ruhm, and J. Waldfogel (2007). “Does prekindergarten improve school preparation and performance?” Economics of Education Review, v.26, 33-51.

[20] NEWEY, Whitney K., James Powell and Francis Vella (1999). ″Nonparametric Estimation of Triangular Simultaneous Equations Models″, *Econometrica*, 67(3): 565-603.

[21] TEMPLE, J. and A. Reynolds (2007). "Benefits and cost of investments in preschool education: evidence from the Child-Parent Centers and related programs". Economics of Education Review, v.26, 126-144.

[22] TEMPLE, J., A. Reynolds, and W. Miedel (2000). "Can early intervention prevent high school dropout?: Evidence from the Chicago Child-Parent Centers". Urban Education, v.35, N.1, 31-56.

9 Tables and Graphs

Table 1: Mean test scores by grade and first school experience

First school experience	Mathematics			Language		Sample Frequency		
	4th grade	8th grade	11th grade	4th grade	8th grade	4th grade	8th grade	11th grade
Daycare centers	197,7	255,7	292,3	188,2	244,8	30,9	32,8	27,9
Kindergarten	187,9	240,7	273,0	176,8	233,9	48,1	49,5	51,4
1st grade elementary	162,6	225,1	252,9	156,4	219,3	18,8	17,1	19,7
2nd grade elementary	144,1	209,1	229,9	134,2	198,8	2,3	0,6	0,9
Total	184,3	241,8	273,5	174,2	233,7	100	100	100
Standard Deviation	47,7	51,6	60,1	45,2	49,9			

Source: SAEB 2005

Table 2: Family characteristics of former DC attendees and non-DC attendees at the 4th grade

	Proportion of DC attendees	Proportion of ECS attendees	Mean score of DC attendees (A)	Mean score of non-DC attendees (B)	DC differential (in sd's, [A-B]/sd)	Sample frequency
Overall	30,9	78,9	197,7	179,5	0,38	100
Mother´s schooling						
Illiterated	16,3	59,4	151,4	159,5	-0,17	5,1
Some elementary (up to 4y)	18,8	72,4	176,0	179,0	-0,06	36,3
Some junior high (5 to 8y)	27,6	79,8	184,2	187,6	-0,07	21,7
Some high-schoool (9 to 11y)	39,0	88,1	208,0	204,5	0,07	16,4
Some college (+12y)	55,1	90,0	222,0	215,5	0,14	20,5
Father´s schooling						
Illiterated	18,5	64,1	156,0	160,7	-0,10	7,8
Some elementary (up to 4y)	20,5	72,5	182,2	181,4	0,02	31,9
Some junior high (5 to 8y)	28,7	78,5	189,8	189,6	0,00	21,9
Some high-schoool (9 to 11y)	35,4	86,8	207,7	205,2	0,05	15,3
Some college (+12y)	54,2	89,9	217,5	211,5	0,13	23,1
Mother at home						
no	29,8	74,1	183,4	182,9	0,01	8,3
yes	30,9	79,4	199,4	192,8	0,14	91,7
Father at home						
no	31,6	77,0	193,5	188,4	0,11	29,0
yes	30,4	79,4	200,7	194,0	0,14	71,0
Parental attitude towards children´s learning						
Parents attend school meetings						
no	27,8	67,9	190,7	178,9	0,25	10,6
yes	31,0	79,8	198,4	192,8	0,12	89,4
Parents motivate children´s studies						
no	25,8	61,7	175,1	174,9	0,00	5,4
yes	30,9	79,4	198,5	192,3	0,13	94,6
Parent's incentivate children to do homework						
no	26,4	67,8	193,9	186,4	0,16	6,4
yes	30,9	79,3	197,9	191,9	0,12	93,6
Parents incentivate reading at home						
no	25,7	73,1	197,4	189,2	0,17	23,3
yes	32,1	80,3	197,8	192,3	0,11	76,7
Parents advise children not to miss classes						
no	30,0	77,0	192,9	182,6	0,22	8,1
yes	30,7	79,1	198,1	192,3	0,12	91,9

Source: SAEB 2005

Table 3: Household characteristics of former DC attendees and non-DC attendees

	Proportion of DC attendees	Proportion of ECS attendees	Mean score of DC attendees (A)	Mean score of non-DC attendees (B)	DC differential (in sd's, [A-B]/sd)	Sample frequency
Number of members						
1-2	31,0	76,2	199,0	187,8	0,23	7,1
3	38,2	86,3	216,7	211,2	0,12	9,2
4-5	37,5	85,4	212,6	207,0	0,12	19,7
6-7	30,0	79,5	190,8	187,3	0,07	41,3
8+	22,9	69,0	179,3	175,5	0,08	22,7
Household structure						
Number of rooms						
0 - 1	24,5	70,3	171,6	171,9	-0,01	17,6
2	30,3	78,9	195,3	191,0	0,09	41,0
3 +	33,6	82,0	207,7	199,3	0,17	41,4
Separate bathroom?						
no	18,4	65,3	169,4	167,1	0,05	11,9
yes	32,3	80,5	199,8	194,3	0,12	88,1
Household goods ans services						
TV						
	18,9	64,9	166,8	167,1	-0,01	12,1
	32,3	80,6	200,1	194,4	0,12	87,9
Radio						
	24,7	71,8	181,6	176,8	0,10	12,7
	31,5	79,7	199,5	193,6	0,12	87,3
Video						
	23,8	74,3	182,1	181,6	0,01	62,7
	42,1	86,1	212,4	206,1	0,13	37,3
Refrigerator						
	16,5	61,8	163,9	161,6	0,05	10,8
	32,4	80,7	199,7	194,4	0,11	89,2
Freezer						
	28,8	49,3	194,3	189,1	0,11	77,5
	37,1	43,7	206,7	200,3	0,13	22,5
Laundry machine						
	25,5	70,2	175,7	172,0	0,08	34,1
	33,3	83,1	206,3	200,2	0,13	65,9
Vacuumer						
	27,7	76,9	190,7	186,5	0,09	81,8
	44,2	87,0	217,4	212,2	0,11	18,2
Car						
	25,0	73,3	181,5	180,6	0,02	58,2
	38,6	86,2	212,3	204,8	0,16	41,8
Computer						
	25,2	75,5	184,1	183,2	0,02	78,4
	50,4	90,3	222,1	217,3	0,10	21,7
Housekeeper						
	26,9	77,3	189,8	187,5	0,05	85,0
	52,0	86,4	220,8	212,9	0,16	15,0

Source: SAEB 2005

Table 4.1: Elementary schools of former DC attendees and non-DC attendees
Pedagogical plan and infrastructure

	Proportion of DC attendees	Proportion of ECS attendees	Mean score of DC attendees (A)	Mean score of non-DC attendees (B)	DC differential (in sd's, [A-B]/sd)	Sample frequency
Pedagogical plan						
Ask homework						
no	29,2	74,5	200,1	185,8	0,30	3,3
yes	30,7	78,8	197,7	191,8	0,12	96,7
Has program against dropout						
no	34,0	46,0	203,7	196,1	0,16	61,4
yes	25,4	51,2	184,9	184,3	0,01	38,6
Has program against repetition						
no	32,8	78,2	206,8	195,9	0,23	37,4
yes	29,4	79,0	191,6	189,2	0,05	62,6
Has extra classes for delayed students						
no	29,3	75,3	187,6	181,0	0,14	30,9
yes	31,3	80,3	202,0	196,2	0,12	69,1
Use of textbook						
no	41,4	81,4	213,1	203,3	0,21	24,9
yes	27,1	77,8	189,9	187,7	0,05	75,1
Infrastructure						
Computers per student						
0 - 0.25	36,5	82,6	204,3	198,2	0,13	64,3
0.25 - 0.5	39,2	85,7	210,4	203,6	0,14	23,5
0.5 - 0.75	37,6	82,3	232,5	217,2	0,32	9,8
0.75 - 1	58,1	94,7	232,3	230,1	0,04	1,3
1+	70,3	95,5	239,1	237,0	0,04	1,1
Computers w/ internet per student						
0 - 0.25	38,3	84,1	207,2	201,3	0,12	69,4
0.25 - 0.5	45,5	85,9	215,0	209,1	0,12	18,9
0.5 - 0.75	41,3	83,1	241,5	227,3	0,30	8,9
0.75 - 1	63,2	94,9	233,1	230,5	0,05	1,3
1+	61,1	89,2	232,0	230,4	0,03	1,5
Ownership						
Public	25,2	76,2	184,6	184,7	0,00	87,5
Private	68,0	95,7	231,2	230,1	0,02	12,5
Class size						
2 - 15	43,1	84,2	214,2	201,7	0,26	3,7
16 - 30	28,6	78,7	198,1	190,3	0,16	47,4
31 - 40	32,6	79,4	197,7	194,3	0,07	43,0
41 - 60	24,7	70,4	172,0	172,9	-0,02	5,9
61 - 90	38,3	76,1	195,2	174,0	0,44	0,0

Source: SAEB 2005

Table 4.2: Elementary schools of former DC attendees and non-DC attendees

Human capital

	Proportion of DC attendees	Proportion of ECS attendees	Mean score of DC attendees (A)	Mean score of non-DC attendees (B)	DC differential (in sd's, [A-B]/sd)	Sample frequency
Human capital						
Principal's experience (years)						
0 - 2	22,3	52,4	184,2	186,0	-0,04	0,7
3 - 5	17,0	62,6	187,7	161,9	0,54	1,7
6 - 10	24,4	70,8	177,7	173,1	0,10	13,2
10+	31,9	80,7	200,6	194,8	0,12	84,4
Principal's tenure on the job (years)						
0 - 2	25,9	75,6	184,6	183,0	0,03	44,8
3 - 5	26,5	78,1	194,5	189,2	0,11	21,9
6 - 10	35,8	81,8	208,4	202,0	0,13	22,7
10+	47,3	88,0	215,0	206,8	0,17	10,6
Principal's education						
Junior high	34,8	67,9	159,6	169,2	-0,20	0,2
High school	22,5	68,2	176,6	169,0	0,16	9,4
College +	31,2	80,0	199,4	193,8	0,12	90,4
Teacher's experience (years)						
0 - 2	24,0	65,7	182,5	172,6	0,21	4,9
3 - 5	28,7	74,7	191,7	187,1	0,09	9,3
6 - 10	31,1	79,0	193,8	192,0	0,04	17,5
10+	31,4	80,2	201,7	194,1	0,16	68,3
Teacher's tenure is that class (years)						
0 - 2	29,1	78,1	190,2	183,9	0,13	4,7
3 - 5	30,9	77,5	182,2	181,7	0,01	9,2
6 +	31,1	79,3	200,7	194,1	0,14	86,1
Teacher's education						
Junior high	36,3	70,2	143,9	152,8	-0,19	0,0
High school	27,5	73,9	183,7	178,8	0,10	26,0
College +	32,0	80,6	203,1	197,0	0,13	74,0

Source: SAEB 2005

Table 5: OLS regressions: coefficients of attendance to daycare centers with Math and language test scores as the dependent variables

Specification	Math			Language		
	R2	Coefficient	St. Error	R2	Coefficient	St. Error
No controls	0,03	17,02	0,08	0,03	18,39	0,08
Individual Attributes (I)	0,11	13,56	0,08	0,14	14,39	0,08
Family						
(I) + Family structure	0,18	8,62	0,12	0,20	7,62	0,12
(I) + Family structure II (II)	0,19	8,88	0,12	0,21	7,88	0,12
(II) + Region (III)	0,23	11,25	0,12	0,24	9,70	0,12
(III) + Family socioeconomic (IV)	0,25	8,87	0,12	0,24	8,23	0,12
(IV) + Grade repetition (V)	0,27	8,42	0,12	0,29	8,01	0,11
School Characteristics						
(V) + Pedagogical Plan (VI)	0,28	8,31	0,12	0,30	7,62	0,11
(VI) + choice of principal (VII)	0,31	5,60	0,12	0,33	4,11	0,11
(VII) + Infrastructure (VIII)	0,33	6,73	0,19	0,36	0,67	0,17
(VIII) + Principal's Traits (IX)	0,39	7,65	0,21	0,41	2,09	0,20
(IX) + Teacher´s Traits	0,42	10,65	0,23	0,43	1,67	0,21
(V) + School Fixed Effects (VI')		3,15	0,75		3,47	0,74
(VI') + Teacher's Traits		3,68	0,84		3,63	0,80

Individual attributes: Sex, Race, Age, Age^2

Family structure: mother living in the household, father living in the household, mother's education, father's education

Family structure II: same as above, replacing continuous parental education by dummy variables of educational level

Region: dummies for Northeast, Central, South and Southeast (omitted: North)

Family - socioeconomic: principal component index of parental attitudes towards learning and socioeconomic index of wealth

Grade repetition: dummy variable indicating repetition

Pedagogical plan: existence of specific programs against dropouts and grade repetition, and extra classes for delayed students

Choice of principal: elected, nominated, promoted

Infrastructure: computers per student, computers w/ internet per student, class size

Principal's traits: education, experience as a director, tenure on the job, sex, race, aeg

Teacher's traits: education, experience as a teacher, experience in this class, sex, race, use of textbook, age

Table 6: OLS regressions: coefficients of attendance to daycare centers and kindergarten, with Math test scores as the dependent variable

Specification	R2	Daycare Centers		Kindergarten	
		Coefficient	St. Error	Coefficient	St. Error
No controls	0,07	8,25	0,09	25,15	0,10
Individual Attributes (I)	0,13	7,15	0,09	20,02	0,10
Family					
(I) + Family structure	0,19	3,74	0,13	16,34	0,14
(I) + Family structure II (II)	0,20	4,04	0,13	15,92	0,14
(II) + Region (III)	0,24	6,93	0,13	13,97	0,14
(III) + Family socioeconomic (IV)	0,26	4,82	0,13	13,76	0,14
(IV) + Grade repetition (V)	0,28	4,69	0,13	12,74	0,14
School Characteristics					
(V) + Pedagogical Plan (VI)	0,29	4,57	0,13	12,76	0,14
(VI) + choice of principal (VII)	0,32	1,95	0,13	12,54	0,14
(VII) + Infrastructure (VIII)	0,34	2,58	0,20	15,51	0,25
(VIII) + Principal's Traits (IX)	0,40	3,66	0,22	14,97	0,27
(IX) + Teacher´s Traits	0,43	6,36	0,25	15,34	0,30
(V) + School Fixed Effects (VI')		0,43	0,80	8,80	0,95
(VI') + Teacher's Traits		1,11	0,90	8,33	1,06

Individual attributes: Sex, Race, Age, Age^2

Family structure: mother living in the household, father living in the household, mother's education, father's education

Family structure II: same as above, replacing continuous parental education by dummy variables of educational level

Region: dummies for Northeast, Central, South and Southeast (omitted: North)

Family - socioeconomic: principal component index of parental attitudes towards learning and socioeconomic index of wealth

Grade repetition: dummy variable indicating repetition

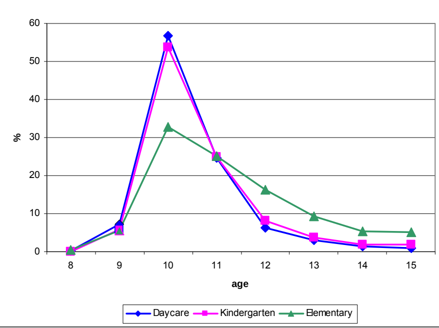
Pedagogical plan: existence of specific programs against dropouts and grade repetition, and extra classes for delayed students

Choice of principal: elected, nominated, promoted

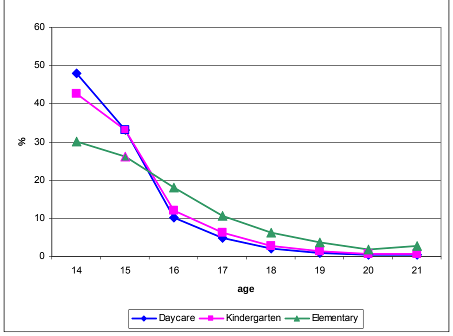
Infrastructure: computers per student, computers w/ internet per student, class size

Principal's traits: education, experience as a director, tenure on the job, sex, race, aeg

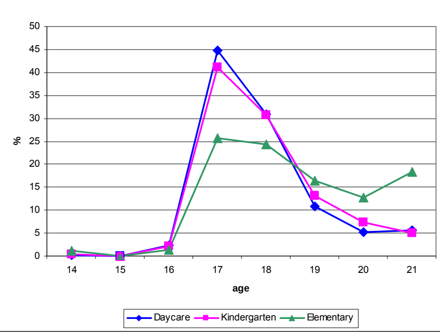
Teacher's traits: education, experience as a teacher, experience in this class, sex, race, use of textbook, age



Graph1.1: Age distribution at the last grade of elementary school, by first educational experience



Graph1.2: Age distribution at the last grade of junior high school, by first educational experience



Graph1.3: Age distribution at the last grade of high school, by first educational experience

Table 7: Probit regressions: coefficients of attendance to daycare centers and kindergarten with an indicator of being in the 4th grade at the correct age as the dependent variable

Specification	Daycare only			Daycare			Kindergarten		
	dy/dx	Coefficient	St. Error	dy/dx	Coefficient	St. Error	dy/dx	Coefficient	St. Error
No controls	0,12	0,46	0,00	0,05	0,18	0,00	0,22	0,71	0,00
Individual Attributes	0,10	0,41	0,00	0,04	0,17	0,00	0,19	0,64	0,00
Family									
(I) + Family structure	0,05	0,22	0,00	0,00	0,01	0,00	0,15	0,58	0,00
(I) + Family structure II (II)	0,05	0,24	0,00	0,01	0,04	0,00	0,15	0,57	0,00
(II) + Region (III)	0,06	0,30	0,00	0,03	0,12	0,00	0,12	0,49	0,00
(III) + Family socioeconomic (IV)	0,05	0,26	0,00	0,02	0,09	0,00	0,11	0,47	0,00
School Characteristics									
(V) + Pedagogical Plan (VI)	0,05	0,26	0,00	0,02	0,09	0,00	0,11	0,47	0,00
(VI) + choice of principal (VII)	0,04	0,21	0,00	0,01	0,04	0,00	0,11	0,47	0,00
(VII) + Infrastructure (VIII)	0,03	0,23	0,00	0,01	0,11	0,00	0,05	0,37	0,00
(VIII) + Principal's Traits (IX)	0,03	0,24	0,00	0,01	0,09	0,00	0,06	0,44	0,00
(IX) + Teacher´s Traits	0,02	0,15	0,00	-0.00*	-0.00*	0,00	0,07	0,48	0,00

Source: SAEB 2005. * Not significant at 10% (all the others are significant at 1%)

Individual attributes: Sex, Race

Family structure: mother living in the household, father living in the household, mother's education, father's education

Family structure II: same as above, replacing continuous parental education by dummy variables of educational level

Region: dummies for Northeast, Central, South and Southeast (omitted: North)

Family - socioeconomic: principal component index of parental attitudes towards learning and socioeconomic index of wealth

Pedagogical plan: existence of specific programs against dropouts and grade repetition, and extra classes for delayed students

Choice of principal: elected, nominated, promoted

Infrastructure: computers per student, computers w/ internet per student, class size

Principal's traits: education, experience as a director, tenure on the job, sex, race, aeg

Teacher's traits: education, experience as a teacher, experience in this class, sex, race, use of textbook, age

Table 8: Probability of entering school, 2005

VARIABLES	Probit regressions		Linear regressions	
	(1)	(2)	(1)	(2)
	Daycare	First grade or more	Daycare	First grade or more
Male	-0,0322	0,0832***	-0,0105	0,0173***
	(0,0256)	(0,028)	(0,0201)	(0,007)
White	-0,0455*	0,009	-0,0142	0,0051
	(0,0271)	(0,03)	(0,0214)	(0,0074)
Age	0,489***	0,013	0,1690*	-0,0373
	(0,16004)	(0,162)	(0,0912)	(0,0444)
Age2	-0,023825***	0,0066	-0,0080**	0,0042***
	(0,0071)	(0,0069)	(0,0038)	(0,002)
Mother's education	0,0311***	-0,0236***	0,0103***	-0,0067***
	(0,0038)	(0,004)	(0,0026)	(0,0011)
Father's education	0,032***	-0,017***	0,0105***	-0,0047***
	(0,0035)	(0,0038)	(0,0024)	(0,001)
Socio-economic Index	1,51***	-1,3353***	0,4797***	-0,3325***
	(0,1608)	(0,1799)	(0,1227)	(0,046)
Private School	0,6175***	-0,3651***	0,2353***	-0,0648***
	(0,0388)	(0,0471)	(0,0338)	(0,0104)
Region	-0,1226***	0,0315**	-0,0382***	0,008***
	(0,0115)	(0,0127)	(0,008)	(0,0033)
Principal's age	0,0014	-0,0018	0,0005	-0,0003
	(0,002)	(0,0022)	(0,0016)	(0,0006)
Principal's education	-0,0136	-0,0166	-0,0040	-0,005
	(0,0122)	(0,0129)	(0,009)	(0,0037)
Principal's salary	-0,000001	0,000001	-0,000003	0,00001
	(0,00002)	(0,00002)	(0,00002)	(0,00001)
Principal's experience	0,0074	-0,0022	0,0021	-0,001
	(0,0051)	(0,0054)	(0,0038)	(0,0014)
Principal's tenure	0,0054**	-0,0098	0,0018	-0,0021***
	(0,0031)	(0,0035)	(0,0025)	(0,0008)
Computers per student	0,0037*	-0,0001**	0,0013	0,0003
	(0,0017)	(0,0019)	(0,0014)	(0,0005)
Teacher's race	-0,0456***	0,0052	-0,0134	0,0022
	(0,0278)	(0,0299)	(0,0213)	(0,0075)
Teacher's age	0,0096***	0,0006	0,003*	0,0001
	(0,0021)	(0,0023)	(0,0016)	(0,0006)
Teacher's education	0,0246	-0,0045	0,008	-0,0013
	(0,0077)	(0,0083)	(0,0056)	(0,0022)
Teacher's salary	-0,00001	-0,00002	-0,00001	-0,00001
	(0,00002)	(0,00003)	(0,00002)	(0,00001)
Teacher's experience	-0,0006	-0,007**	-0,0003	-0,0018***
	(0,0031)	(0,0034)	(0,00002)	(0,0009)

Table 8: Probability of entering school, 2005 (Cont)				
	Probit regressions		Linear regressions	
VARIABLES	(1)	(2)	(1)	(2)
	Daycare	First grade or more	Daycare	First grade or more
Number of students in the classrooom	-0,0015	0,0063***	-0,0006	0,0014***
	(0,0018)	(0,002)	(0,0015)	(0,0005)
Number of unused classrooms 1997	0,8569**		0,2618	
	(0,3771)		(0,2614)	
Number of unused classrooms 1998	2,928***	-0,2707	0,8829***	-0,1232
	(0,3907)	(0,4317)	(0,3007)	(0,1044)
Number of unused classrooms 1999		-0,8269		-0,2159
		(0,6059)		(0,1552)
Number of unused classrooms 2000		0,194		0,0288
		(0,6456)		(0,1714)
Number of unused classrooms 2001		-1,5645**		-0,4223***
		(0,546)		(0,1357)
Hospitalization by repiratory diseases_0a3_1995	0,0063***		0,0018	
	(0,0017)		(0,0012)	
Hospitalization by repiratory diseases_0a3_1996	-0,0033*		-0,001	
	(0,0017)		(0,0011)	
Hospitalization by repiratory diseases_0a3_1997	0,0008		0,0003	
	(0,0016)		(0,0011)	
Hospitalization by repiratory diseases_0a3_1998	-0,00005		-0,00002	
	(0,0013)		(0,0001)	
Hospitalization by digestive diseases_0a3_1995	-0,0117***		-0,0035**	
	(0,0024)		(0,0016)	
Hospitalization by digestive diseases_0a3_1996	0,0101***		0,0028*	
	(0,0026)		(0,0016)	
Hospitalization by digestive diseases_0a3_1997	-0,0017		-0,0004	
	(0,0026)		(0,0015)	
Hospitalization by repiratory diseases_0a3_1998	-0,0031		-0,001	
	(0,0019)		(0,0012)	
Hospitalization by repiratory diseases_4a6_1998		0,0015		0,0005
		(0,0026)		(0,0007)
Hospitalization by repiratory diseases_4a6_1999		-0,0043		-0,0013
		(0,0034)		(0,0009)
Hospitalization by repiratory diseases_4a6_2000		-0,0031		-0,0006
		(0,0035)		(0,0009)
Hospitalization by repiratory diseases_4a6_2001		0,004		0,001
		(0,0033)		(0,0009)
Hospitalization by digestive diseases_4a6_1998		0,0042		0,0014
		(0,0034)		(0,001)
Hospitalization by digestive diseases_4a6_1999		-0,0069		-0,0019*
		(0,0039)		(0,0011)
Hospitalization by digestive diseases_4a6_2000		0,003		0,0008
		(0,0046)		(0,0013)
Hospitalization by digestive diseases_4a6_2001		0,0032		0,0006
		(0,0044)		(0,0012)
	-4,6615***	-0,4915	-1,1427	0,4575*
Constant	(0,9295)	(0,9697)	(0,5647)	(0,2592)
Test of joint significance of the observables[1]	4,1200	1,006	0,7375	1,3595
P-value	0,0000	0,4452	0,6898	0,1775
Observations	11779	11779	11779	11779
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				
[1]LR and F for the probit and linear regressions, respectively				

Table 9: Test Scores, 2005			
VARIABLES	OLS	CF	IV
Male	-13,177***	-13,093***	-12,454***
	(0,735)	(0,715)	(1,779)
White	2,828***	2,699***	2,057
	(0,784)	(0,761)	(1,86)
Age	-13,315***	-12,677***	-4,372
	(4,025)	(3,9)	(9,425)
Age2	0,363**	0,335**	0,156
	(0,173)	(0,168)	(0,411)
Mother's education	0,428***	0,436***	0,551*
	(0,109)	(0,106)	(0,32)
Father's education	0,404***	0,417***	0,732*
	(0,101)	(0,098)	(0,303)
Socio-economic Index	36,055***	36,151***	39,124***
	(4,754)	(4,573)	(14,96)
Type of School	21,901***	21,922***	33,048***
	(1,24)	(1,185)	(5,253)
Region	1,087***	1,048***	-0,66
	(0,293)	(0,285)	(0,855)
Principal's age	0,048	0,053	0,048
	(0,058)	(0,056)	(0,137)
Principal's education	0,363	0,359	-0,339
	(0,339)	(0,335)	(0,835)
Principal's salary	0,005***	0,005***	0,005
	(0,001)	(0,001)	(0,001)
Principal's experience	0,175	0,181	0,229
	(0,147)	(0,142)	(0,33)
Principal's tenure	0,105	0,096	0,107
	(0,09)	(0,087)	(0,224)
Computers per student	0,292***	0,291***	0,413
	(0,05)	(0,048)	(0,121)
Teacher's race	1,734**	1,648**	0,725
	(0,776)	(0,75)	(1,853)
Teacher's age	0,012	0,021	0,233
	(0,05)	(0,059)	(0,153)
Teacher's education	0,222	0,243	0,664
	(0,22)	(0,214)	(0,521)
Teacher's salary	0,002***	0,002***	0,002***
	(0,001)	(0,001)	(0,0017)
Teacher's experience	0,244***	0,235***	0,056
	(0,089)	(0,087)	(0,213)
Number of students in the classrooom	-0,157***	-0,144***	-0,072***
	(0,052)	(0,05)	(0,129)
Daycare	14,24***	12,793***	68,412***
	(1,514)	(1,1731)	(19,122)
Kintergarten	12,893***	11,106***	-100,1***
	(1,68)	(1,212)	(26,084)
Control Function		-3,074***	
		(1,164)	
Constant	210,35***	209,45***	180,39***
	(24,151)	(23,278)	58,163
Observations	11779	11779	11779
R-squared	0.352	0.366	0.356
Test: Daycare=Preschool			
F-statistics	0,0767	6,8810	0,8541
P-value	0,7818	0,0089	0,3554
Robust standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Table 10: Probability of entering school, 2005-2007

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Daycare	Daycare	Daycare	Daycare	Kintergarten	Kintergarten
Male	0.0382	0.0366	0.0431	-0.00195	0.0412	-0.00429
White	0.00355	0.00406	0.00339	0.00155	0.00398	0.00229
Age	-0.0317**	-0.0327**	-0.0296**	-0.0160	-0.0310**	-0.0176
Mother's education	0.0125***	0.0127***	0.0125***	-0.00154	0.0127***	-0.00130
Father's education	0.0128***	0.0130***	0.0129***	-0.00631*	0.0131***	-0.00606*
Grade repetition	0.0112	0.0124	0.0139	-0.0264	0.0152	-0.0248
Socio-Economic Index	0.0170	0.0185	0.0187	-0.0471***	0.0205	-0.0449***
Principal's age	-0.000743	-0.000708	-0.000763	0.00115	-0.000723	0.00120
Principal's education	0.00722	0.00712	0.00807	0.00761	0.00793	0.00744
Principal's salary	-6.38e-06	-6.00e-06	-4.90e-06	5.08e-07	-4.45e-06	1.07e-06
Principal's experience	0.00410*	0.00408*	0.00375*	-0.00592***	0.00373*	-0.00594***
Principal's tenure	-0.000274	-0.000244	-0.000194	0.000456	-0.000158	0.000500
computer per student	-0.000991	-0.000959	-0.000940	0.00177*	-0.000903	0.00181**
Teacher's race	-0.00202	-0.00122	-0.000205	0.00844	0.000738	0.00961
Teacher's age	-7.99e-05	-0.000101	-5.29e-05	0.000438	-7.79e-05	0.000407
Teacher's education	0.00460	0.00444	0.00410	0.000981	0.00391	0.000744
Teacher's salary	1.01e-05	1.01e-05	8.62e-06	5.11e-08	8.56e-06	-2.50e-08
Teacher's experience	0.000394	0.000440	0.000410	-0.000433	0.000469	-0.000360
Number of students in the classrooom	0.000614	0.000603	0.000557	8.17e-05	0.000542	6.40e-05
Teacher assigns homework	-0.0553	-0.0391	-0.0711	-0.0484	-0.0509	-0.0234
Parental attitude towards children´s learning	0.0238		0.0295	0.0365		
Number of unused classrooms (99-97)	-0.138	-0.143	-0.245	0.0467	-0.251	0.0400
Number of unused classrooms (98-00)	0.110	0.116	0.138	-0.259	0.147	-0.248
Number of unused classrooms (01-99)			-0.109	-0.0139	-0.110	-0.0143
Number of unused classrooms (02-00)			0.117	-0.167	0.120	-0.164
Number of unused classrooms (03-01)			-0.178	-0.0761	-0.183	-0.0820
Hospitalization by repiratory diseases_0a3_9795	0.000940	0.000956	0.00101	-0.000866	0.00103	-0.000850
Hospitalization by repiratory diseases_0a3_9896	-0.000103	-0.000101	0.000274	-0.000752	0.000280	-0.000745
Hospitalization by repiratory diseases_0a3_9997	0.000514	0.000509	0.000582	-0.000382	0.000556	-0.000414
Hospitalization by repiratory diseases_0a3_0098	0.000917	0.000904	0.000845	-0.00101	0.000832	-0.00103
Hospitalization by digestive diseases_0a3_9795	-0.00142	-0.00145	-0.00201**	0.00195**	-0.00204**	0.00192**
Hospitalization by digestive diseases_0a3_9896	-0.000135	-0.000122	0.000440	0.00107	0.000454	0.00109
Hospitalization by digestive diseases_0a3_9997	-0.00148*	-0.00149*	-0.00274***	0.00131	-0.00272***	0.00132
Hospitalization by digestive diseases_0a3_0098	4.14e-05	5.88e-05	0.000701	-4.53e-05	0.000702	-4.36e-05
Hospitalization by repiratory diseases_4a6_0098			0.000512	-0.000976	0.000544	-0.000937
Hospitalization by repiratory diseases_4a6_0199			0.000524	0.00124	0.000441	0.00113
Hospitalization by repiratory diseases_4a6_0200			-0.00104	-0.000876	-0.00102	-0.000852
Hospitalization by repiratory diseases_4a6_0301			0.00244*	-0.00135	0.00241*	-0.00138
Hospitalization by digestive diseases_4a6_0098			0.000967	-0.00154	0.000990	-0.00151
Hospitalization by digestive diseases_4a6_0199			-0.00576***	0.00506***	-0.00567***	0.00518***
Hospitalization by digestive diseases_4a6_0200			0.00277	-0.00151	0.00273	-0.00155
Hospitalization by digestive diseases_4a6_0301			-0.00275**	0.00318**	-0.00269**	0.00324**
Constant	0.0478***	0.0482***	0.0430***	-0.134***	0.0435***	-0.133***
Observations	832	832	832	832	832	832
R-squared	0.099	0.098	0.115	0.068	0.114	0.067
Test:						
F-statistics	2.83	2.91	2.44	1.37	2.49	1.38
P-value	0.000	0.000	0.000	0.061	0.000	0.058

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

Table 11: Test Scores, 2005-1007

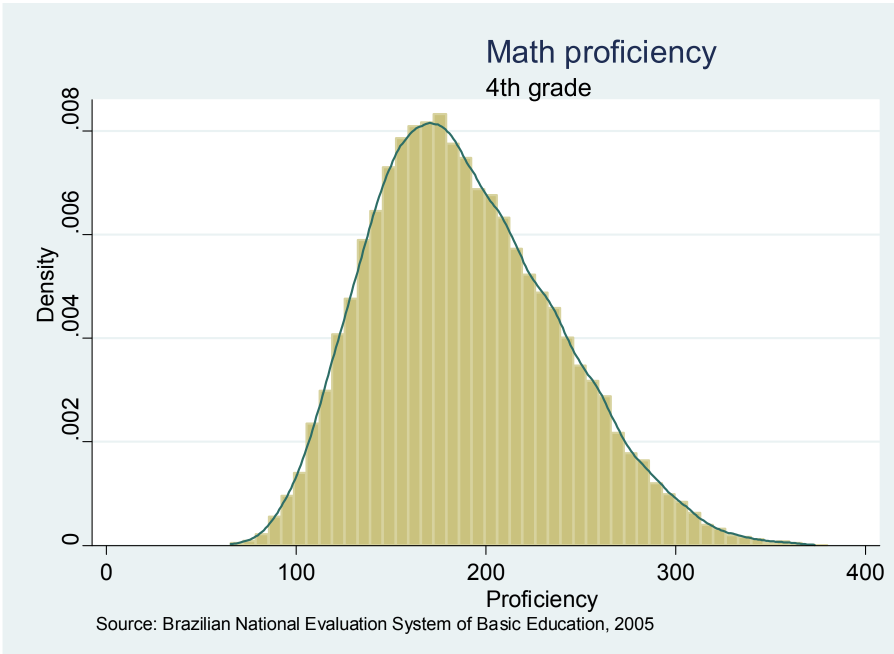
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	-9.275**	-10.23***	-10.88***	-11.93***	-9.789***	-10.73***	-10.92***	-11.88***
	(3.785)	(3.813)	(4.061)	(4.076)	(3.762)	(3.786)	(4.078)	(4.077)
White	1.876	2.180	2.623	2.837	1.887	2.177	2.579	2.786
	(3.159)	(3.188)	(3.265)	(3.295)	(3.137)	(3.163)	(3.246)	(3.275)
Age	-0.120	-0.628	-0.0199	-0.466	0.290	-0.167	0.0785	-0.391
	(1.098)	(1.101)	(1.432)	(1.456)	(1.097)	(1.100)	(1.573)	(1.631)
Mother's education	-0.378	-0.301	-0.486	-0.445	-0.466	-0.399	-0.436	-0.372
	(0.331)	(0.333)	(0.477)	(0.483)	(0.329)	(0.332)	(0.453)	(0.464)
Father's education	0.301	0.372	0.265	0.315	0.270	0.336	0.380	0.459
	(0.290)	(0.292)	(0.454)	(0.460)	(0.288)	(0.290)	(0.384)	(0.393)
Grade repetition	-27.57***	-26.88***	-28.33***	-27.69***	-27.25***	-26.58***	-28.04***	-27.37***
	(3.754)	(3.785)	(3.932)	(3.967)	(3.729)	(3.757)	(3.906)	(3.933)
Socio-Economic Index	-2.303*	-1.498	-2.163	-1.412	-1.875	-1.085	-1.445	-0.620
	(1.261)	(1.257)	(1.427)	(1.443)	(1.258)	(1.253)	(1.533)	(1.498)
Principal's age	0.0325	0.0486	0.0205	0.0398	0.0269	0.0419	0.00353	0.0193
	(0.0779)	(0.0785)	(0.0833)	(0.0837)	(0.0774)	(0.0779)	(0.0812)	(0.0826)
Principal's education	-0.0462	-0.116	-0.271	-0.347	-0.213	-0.289	-0.363	-0.435
	(0.437)	(0.440)	(0.500)	(0.503)	(0.436)	(0.439)	(0.574)	(0.576)
Principal's salary	0.00158*	0.00177**	0.00157*	0.00176**	0.00163*	0.00181**	0.00154*	0.00171**
	(0.000868)	(0.000875)	(0.000905)	(0.000910)	(0.000863)	(0.000869)	(0.000895)	(0.000898)
Principal's experience	0.0666	0.0518	0.0894	0.0742	0.104	0.0918	0.187	0.189
	(0.181)	(0.183)	(0.220)	(0.222)	(0.180)	(0.182)	(0.209)	(0.211)
Principal's tenure	-0.109	-0.0972	-0.144	-0.129	-0.114	-0.103	-0.151	-0.137
	(0.130)	(0.131)	(0.133)	(0.135)	(0.129)	(0.130)	(0.132)	(0.134)
computer per student	0.0980	0.113	0.0840	0.102	0.0820	0.0957	0.0579	0.0710
	(0.0766)	(0.0772)	(0.0831)	(0.0835)	(0.0762)	(0.0767)	(0.0828)	(0.0837)
Teacher's race	2.600*	3.115**	3.226**	3.708***	2.469*	2.950**	3.045**	3.488**
	(1.381)	(1.388)	(1.426)	(1.432)	(1.372)	(1.378)	(1.435)	(1.451)
Teacher's age	-0.0205	-0.0301	-0.0550	-0.0647	-0.0232	-0.0325	-0.0541	-0.0631
	(0.0693)	(0.0699)	(0.0721)	(0.0727)	(0.0688)	(0.0694)	(0.0716)	(0.0721)
Teacher's education	0.00909	-0.0608	0.0797	-0.00598	-0.0334	-0.102	0.0789	0.00259
	(0.289)	(0.291)	(0.320)	(0.321)	(0.288)	(0.289)	(0.323)	(0.322)
Teacher's salary	0.000999	0.000902	0.00110	0.00101	0.000945	0.000850	0.00112	0.00106
	(0.00107)	(0.00108)	(0.00114)	(0.00115)	(0.00107)	(0.00108)	(0.00113)	(0.00114)
Teacher's experience	0.0444	0.0705	0.0390	0.0616	0.0453	0.0701	0.0440	0.0668
	(0.0923)	(0.0929)	(0.0949)	(0.0958)	(0.0917)	(0.0922)	(0.0940)	(0.0947)
Number of students in the classrooom	-0.0367	-0.0432	-0.0302	-0.0379	-0.0447	-0.0514	-0.0296	-0.0358
	(0.0640)	(0.0646)	(0.0683)	(0.0688)	(0.0636)	(0.0641)	(0.0682)	(0.0686)
Teacher assigns homework	12.07***	20.35***	12.01***	20.18***	13.04***	20.95***	12.36**	20.17***
	(4.179)	(3.689)	(4.662)	(4.023)	(4.159)	(3.665)	(4.954)	(4.090)
Parental attitude towards children´s learning	12.47***		12.24***		11.87***		11.81***	
	(3.053)		(3.231)		(3.036)		(3.452)	
Enter school at daycare	2.275	2.769	8.885	11.56	**10.63*** **	**11.61*** **	5.982	6.716
	(2.886)	(2.910)	(27.43)	(27.54)	**(3.709)**	**(3.732)**	(26.28)	(23.71)
Enter school at preschool					**13.03*** **	**13.83*** **	13.85	15.21
					(3.671)	**(3.696)**	(23.41)	(23.71)
residual daycare (CF)			-7.535	-9.934			3.753	3.899
			(26.67)	(26.79)			(25.69)	(26.10)
residual kindergarten (CF)							-0.756	-1.190
							(22.94)	(23.24)
Constant	18.01***	18.25***	17.26***	17.37***	19.45***	19.76***	19.41***	19.81***
	(1.022)	(1.030)	(1.798)	(1.820)	(1.092)	(1.099)	(2.704)	(2.708)
R-squared	0.150	0.133	0.164	0.146	0.163	0.147	0.177	0.161
Observations	861	861	832	832	861	861	832	832
Test: Daycare=Preschool								
F-statistics					0,58	0,49	0,21	0.24
P-value					0,4464	0,4853	0,6483	0.6259

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

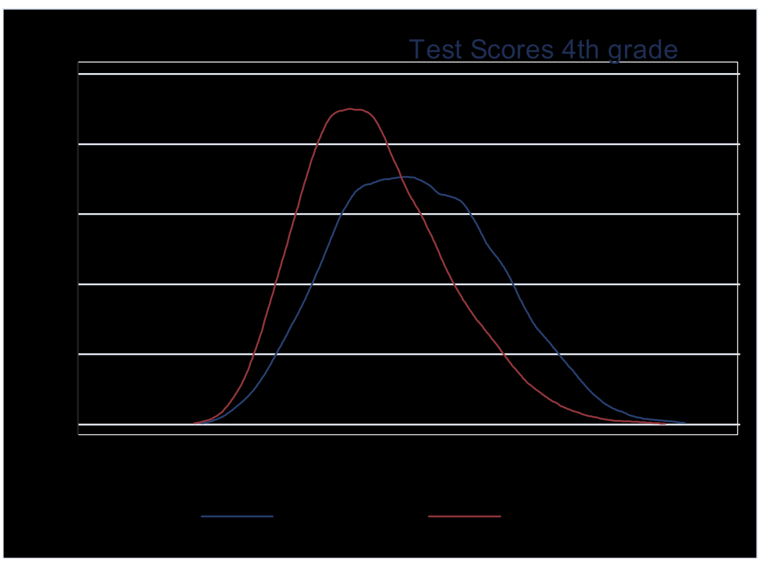
Table A1: Descriptive Analysis

Variables	N	Mean/Proportion	SD	Min.	Max.
Proficiency	41.783	188,9	48,94	65,43	373,44
Student´s Characteristics					
Attended Daycare	41.783	0,38	0,48	0	1
Kindergarten	41.783	0,76	0,42	0	1
Male	40.964	0,50	0,50	0	1
White	39.751	0,37	0,48	0	1
Age	41.170	10,66	1,20	8	15
Age*2	41.170	115,14	27,87	64	225
Mother's schooling	27.386	8,80	4,93	0	15
Father's schooling	23.730	8,91	5,15	0	15
Parents participate in school´s meetings	41.783	0,88	0,31	0	1
Has repeted a grade	41.783	0,26	0,44	0	1
Socio-economic Level	39.919	0,02	0,11	-0,30	0,25
School´s characteristics					
Principal's age	40.706	43,92	7,81	24	55
Principal´s Education	40.194	15,29	1,33	8	17
Principal's salary	38.341	1.748,00	842,55	300,00	3.100,00
Principal´s Experience	40.720	13,47	3,01	1	15
Principal´s Experience as principal	40.309	6,23	5,05	1	15
How the principal is selected					
Election	39.854	0,20	0,40	0	1
Selected by professional outside	39.854	0,14	0,35	0	1
Selected by politicians	39.854	0,12	0,32	0	1
Selected by experts in educations	39.854	0,10	0,29	0	1
Selected by the community	39.854	0,10	0,29	0	1
Selected and elected	39.854	0,13	0,33	0	1
Others	39.854	0,21	0,40	0	1
Public school	41.783	0,70	0,45	0	1
Region					
Middle-West	41.783	0,14	0,34	0	1
North	41.783	0,20	0,40	0	1
Northeast	41.783	0,36	0,47	0	1
South	41.783	0,14	0,34	0	1
Southeast	41.783	0,16	0,36	0	1
Computers in school	35.698	10,94	9,91	3	30
Classroom´s characteristics					
Teacher´s race (white)	37.878	0,45	0,49	0	1
Teacher's age	38.079	39,08	8,86	24	55
Teacher´s salary	36.953	1.022,61	626,2692	300	3.100,00
Teacher's experience	37.759	13,24	6,11	0,5	20
Students have book	41.783	0,62	0,48	0	1
Number the students in the classroom	41.504	29,72	8,02	2	67

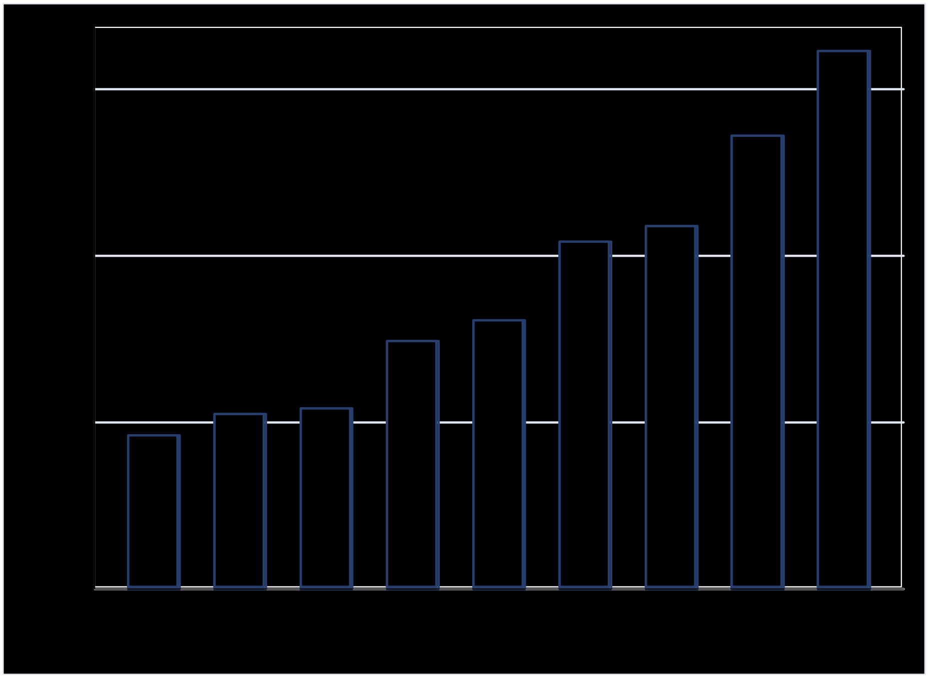
Source: Brazilian National Evaluation System of Basic Education (SAEB), 2005



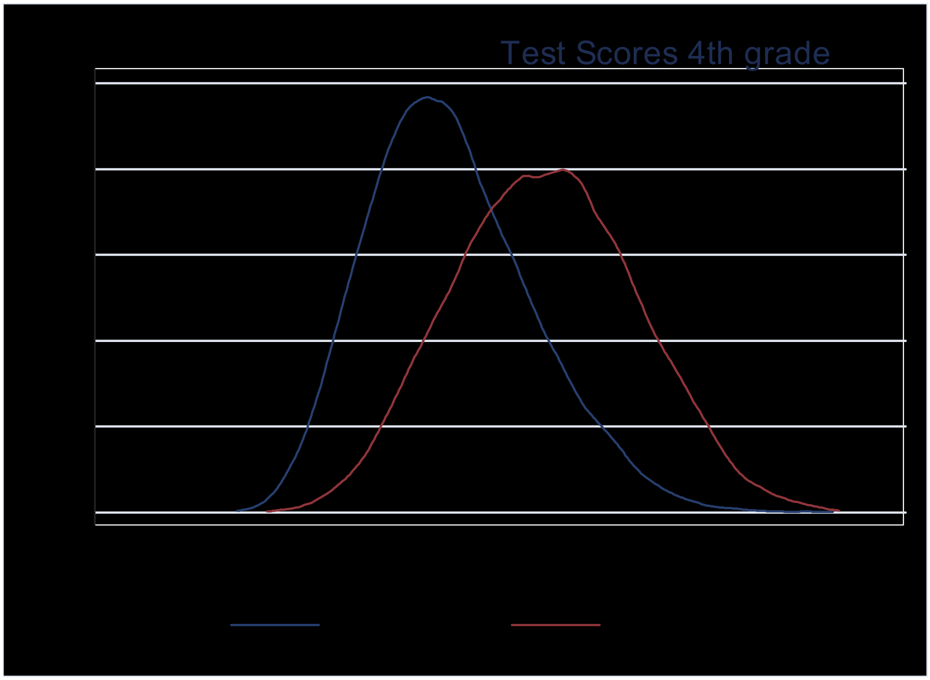
Graph A1



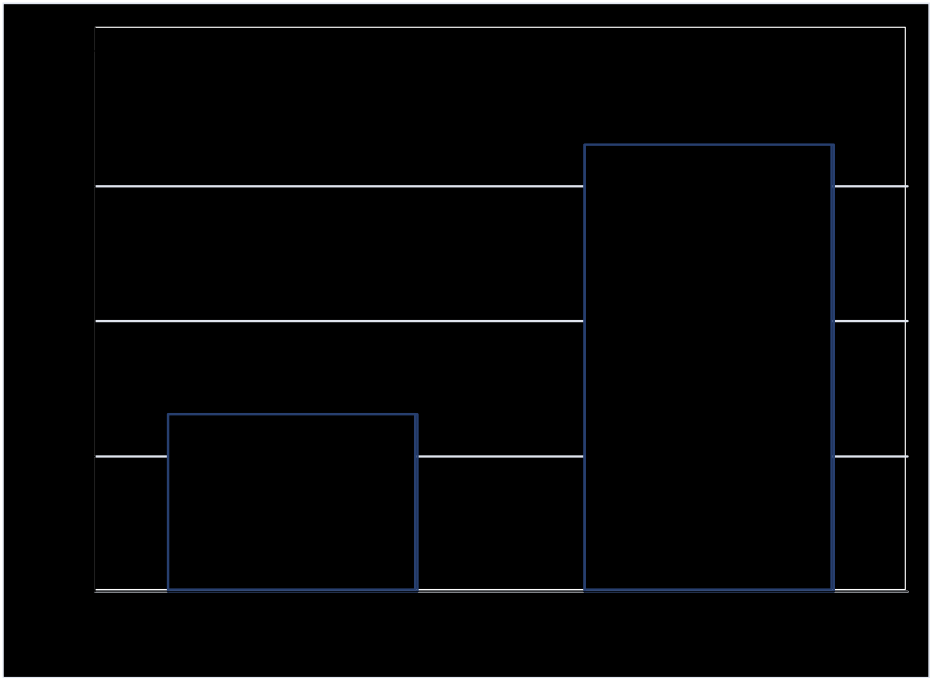
Graph A2



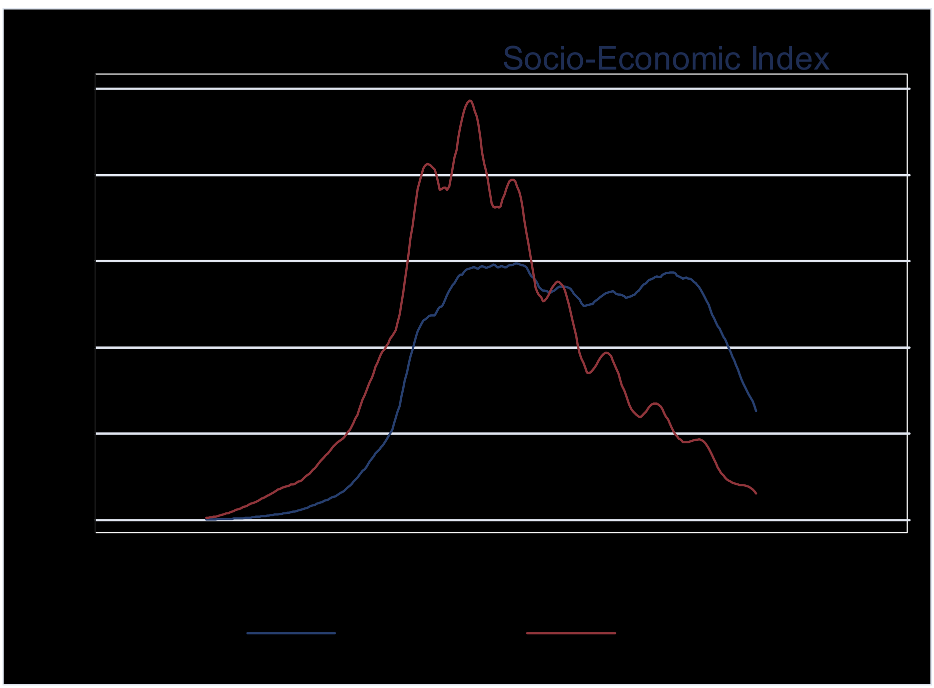
Graph A3



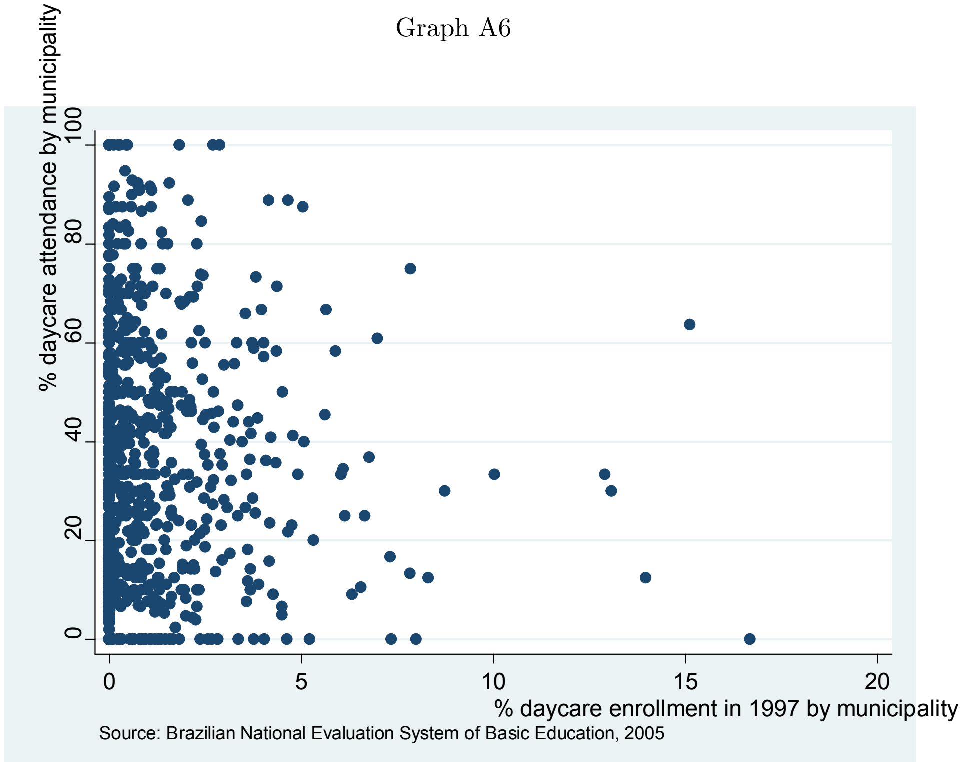
Graph A4



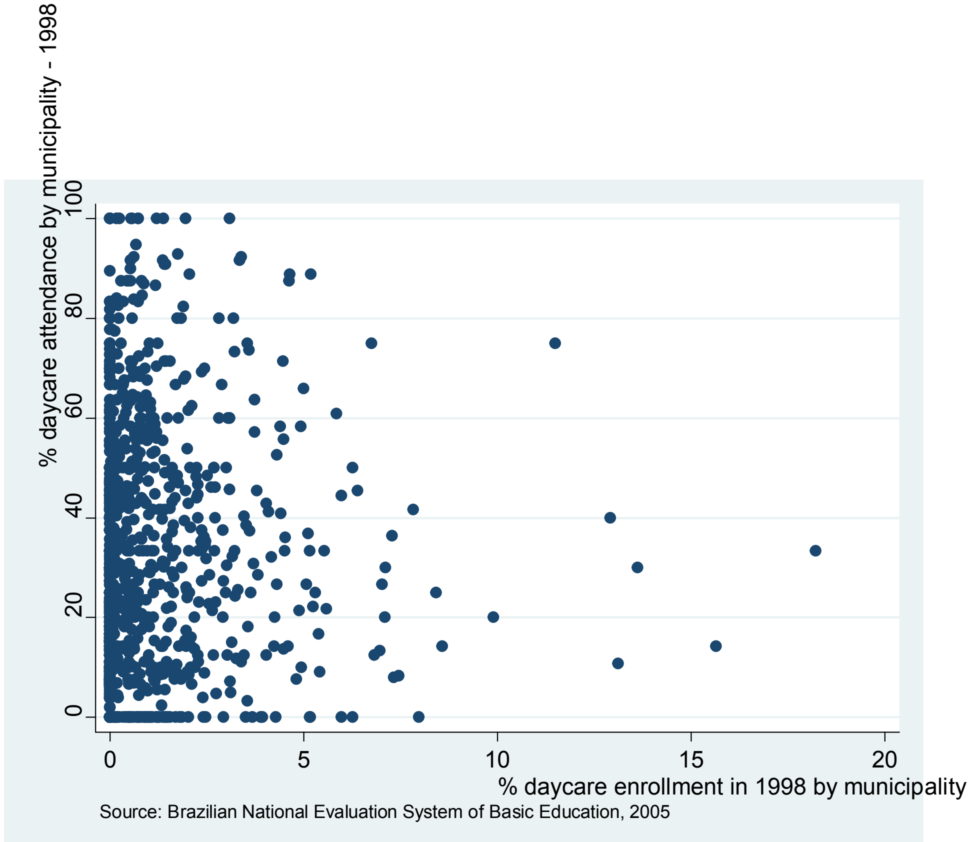
Graph A5



Graph A6



Graph A7


% daycare attendance by municipality - 1998
100
80
60
40
20
0
0
5
10
15
20
% daycare enrollment in 1998 by municipality
Source: Brazilian National Evaluation System of Basic Education, 2005

Graph A8

WORKING PAPER 10/2011 • MAY, 2011 •