The Effect of Preschool on Proficiency Level: An Analysis from Longitudinal data†

Igor Procópio+, Ricardo Freguglia*, Flávia Chein§

RESUMO

Este artigo estima o efeito da pré-escola sobre o nível de proficiência de alunos do primeiro ciclo do ensino fundamental, utilizando um banco de dados longitudinal produzido pelo Projeto GERES. O uso do GERES permite a avaliação da evolução da aprendizagem, bem como o controle de características não observadas dos estudantes. O efeito da pré-escola foi estimado através de uma função de produção educacional considerando o processo dinâmico de aprendizado, conforme Cunha e Heckman (2007) e Todd e Wolpin (2003). A estratégia empírica utilizada foi uma adaptação da estimação por valor adicionado e especificação cumulativa. Os resultados encontrados indicam o efeito cumulativo dos insumos e a presença de heterogeneidade individual. Para Matemática o efeito direto encontrado foi positivo e significativo. Já para português, os coeficientes de pré-escola apresentaram resultados não significativos, indicando que não há efeito direto sobre as habilidades em português. O efeito indireto, manifestado através da proficiência inicial, foi positivo e crescente ao longo dos anos para Matemática. Por outro lado, os estudantes com maiores habilidades em linguagem no início do ensino fundamental aumentam suas diferenças comparativamente aos menos habilidosos, mas esse efeito se inverte ao longo dos anos, indicando que a desigualdade de habilidades em Português podem ser reduzidas durante o período escolar.

ABSTRACT

This article estimates the preschool effect on the child's proficiency level based on a database built by GERES Project, which is a longitudinal study that monitories a sample of students from 1st grade for four years. The use of GERES allows the evaluation of the evolution of learning, controlling for unobserved characteristics of the students that affect proficiency or performance. Following Cunha & Heckman (2007) and Todd & Wolpin (2003), the preschool effect is estimated through an educational production function that incorporates a dynamic process of learning. The empirical strategy is an adaptation of value added and cumulative specification estimates. Our findings show evidence of cumulative effect of inputs and the presence of individual heterogeneity. In the case of Mathematics the direct effect of preschool on child's proficiency was positive and significant, while in Portuguese, there is no direct effect. Furthermore, the indirect effect shown by the coefficient of initial proficiency is positive and increased over the years for Math. On the other had, the students with higher skills in language at the beginning of elementary school expanded their differences in relation to less skilled students, but this effect inverts over the years, indicating that the inequality of skills in Portuguese can be reduced during the schoolar period.

JEL Classification: I21; I24

Key Words: Preschool, GERES, Dynamic Process of Learning, Proficiency, Educational Production Function

Palavras Chaves: Pré-escola, GERES, Dinâmica do Aprendizado, Proficiência, Função de Produção Educacional

Área 12: Economia Social e Demografia Econômica

[0]†We are thankful to CAPES and to INEP for financial support through "Observatório da Educação-2010: Projeto 34 - Análise da Evolução da Educação Básica no Brasil a partir de indicadores de fluxo e proficiência". We are also grateful to CAED/UFJF for providing GERES database.

+PPGEA/UFJF; Email: igor.procopio@ufjf.edu.br

*PPGEA/UFJF; Email: ricardo.freguglia@ufjf.edu.br

§PPGEA/UFJF; Email: flavia.chein@ufjf.edu.br

1 Introduction

Results from the 2009 Program for International Student Assessment (PISA) show that Brazil progressed in relation to the results from the edition of the previous evaluation, done in 2006. However, this progress was not sufficient to remove Brazil from the bottom of the global ranking. From a total of sixty-five countries evaluated in 2009, Brazil occupied the 53rd position in the ranking in reading and the 57th position in the ranking in mathematics. These results show the need for public policies aimed at an improvement in education in Brazil. Among the diverse types of policies in this area, such as valorization and capacity building of teachers, better infrastructure in schools, increasing the school day, the policies directed at younger children, under 6 years of age, stand out.

Graph 1 shows the progress of Brazil during the past decade, in the years in which the international assessment was applied. A monotonic increase for proficiency in mathematics can be seen, going from a mean of 334 points in the year 2000 to 386 points in 2010. As for reading proficiency, despite the increase between the years 2000 and 2010, this growth was less accentuated; there was also a fall between the years 2003 and 2006.

[Insert Graph 1]

In comparison with participating Latin American countries, Brazil appears in front of Argentina and Colombia, but behind Mexico, Uruguay and Chile. China, a developing country and belonging to BRIC, showed the best result among all the countries participating in the 2009 edition of PISA, both in reading and in mathematics. The comparison with these countries shows that, despite the Brazilian progress between 2000 and 2009 being inferior only to the progress of Chile and Luxembourg, in absolute terms, the country still needs to progress much to approach the countries with the best results.

As a strategy for improvements in the level of learning of children the need for investments in early childhood, a period that precedes the entrance of the child into mandatory regular schooling, that is, under six years of age, has come to prominence. Recent advances in neuroscience have demonstrated the importance of investments and stimuli in the first years of live to cerebral development and learning capacity. In this sense, several studies evaluated the effect of interventions done in early childhood on the success of school learning and labor market allocation. Considering this evidence, preschool attendance is identified as a mechanism of stimulating cognitive and non-cognitive development of children, mainly for children in vulnerable situations.

Cunha and Heckman (2007, 2009, 2010) show that educational investments in early childhood focused on children from families of low socioeconomic level present high rates of return in contrast with investments done in adolescence. According to the authors, there is a tradeoff between equality and efficiency in the last kind of investment, but not in the early childhood's investments.

The international literature shows the existence of several programs directed at the promotion of child development, mainly for children of disadvantaged families. In the United States the High/Scope Perry PreSchool Project, the Carolina Abecedarian Project, the Early Training Project, and Head Start stand out. In Latin America, the best-known programs are Mexico's Oportunidades, Bolivia's Proyecto de Desarollo Infantil and the construction of preschools in Argentina.

In Brazil, the analysis of the impact of investments in early childhood focuses on the effect of preschools and day care centers. Enrollment in preschool is not mandatory and public provision is not universal as occurs in elementary school. Data from the World Bank (2006) show that children of families from poorer classes have lower attendance in child education. Considering the evidence of the positive effects on child development in children who attend preschool, this unequal access has a perverse effect on the creation of inequalities.

The Brazilian government is already taking some actions in order to improve children skills. One first step was the reduction of the mandatory age for enrollment in elementary school from 7 to 6 years of age.

Another initiative was the creation of the Fund for the Maintenance and Development of Basic Education and for the Valorization of Educational Professionals (FUNDEB) that serves educational levels from day care centers to secondary schools.

Data from the National Survey by Household Sample (PNAD) show an increase in the percentage of children aged 6 years old or less attending school in the past decade. Graph 2 shows the evolution of attendance for the age group of 0 to 6 years, a period denominated early childhood, and of the specific period of preschool, 4 to 6 years. The results show a monotonic increase of school attendance and that for the preschool period the values go from 65% at the beginning of the decade to a little more than 81% in the year 2009. For early childhood as a whole, attendance was around 35% in 2001 and reached 47% in 2009.

[Insert Graph 2]

A series of recent studies evaluates the effect of preschool on proficiency tests of students in primary and secondary education. The results found indicate a positive relationship between preschool attendance and academic success. This paper contributes to the analysis of the impact of investments in early childhood using a database still little explored in the domestic literature, the 2005 School Generation project – GERES. GERES is a longitudinal study that monitories a sample of students from 1st grade for four years. The use of a longitudinal study allows the evaluation of the evolution of learning, controlling for unobserved characteristics of the students that affect proficiency or performance. Policy education's evaluators recommend such a strategy as the best option for identifying the determinants of learning.

This paper is structured in the following manner. Besides this introduction, Section 2 brings a brief review of the international and domestic literature about the evaluation of the effect of early childhood programs. Section 3 presents the database used and the strategy of identifying the effect of preschool. Section 4 shows some descriptive statistics and Section 5 brings the econometric results. Finally, in Section 6 conclusions are made.

2 Literature Review

Studies of neuroscience have demonstrated the importance of the first years of life on cerebral formation and development (Knudsen, 2004; Casella et al., 2011). Advances in magnetic resonance equipment allow us to see how the brain reacts to stimuli, permitting us to detect what types of stimuli and in what phase of the lifecycle the results are most favorable for cerebral development.

In face of these advancements, in the understanding of cerebral function, Cunha and Heckman (2007) develop an economic model for the formation of skills throughout the lifecycle of an individual. This model presents by means of a production function the learning dynamic, emphasizing the importance of investments done in the first years of life for educational and labor market success. The understanding of the process of formation of individual capacities expands the knowledge about the origins of inequality between people and contributes to the design of policies for inequalities reduction.

Cunha *et al* (2006) and Cunha and Heckman (2007, 2009) show the importance of investment in disadvantaged children and that the return from these investments is high when done in the first years of life. Furthermore, Cunha and Heckman (2007) present evidence and theoretical arguments that there is not a tradeoff between equality and efficiency in investments in early childhood. Public policies whose aims are to improve human capital should be focused on children from poor families, given the lower private investment of these families on the training of their children's skills.

The international literature presents a vast analysis of governmental programs focused on child development. Barnett (1992), Currie (2001), Currie and Blau (2006), Cunha *et al.* (2006) present a review of the principal programs and results found. In general, the results point to a positive relationship between the interventions and gains in skills.

The Unites States stands out for some programs implemented with random selection of participants, seeking to obtain more precise assessments of the results. High/Scope Perry PreSchool was implemented during the years from 1962 to 1967 in about 120 low-income students between 3 and 4 years old. The children were divided into treatment and control groups. The treatment group received half a day of preschool, weekly visits during eight month per year over the course of two years. The program collected information about the children from the two groups of 3 to 15 years and later at 17, 27 and 40 years. The results of the assessment show that the treatment group showed a low repetition rate, high performance in cognitive tests, high rates of elementary school completion, besides lower involvement in crime, high salaries and low use of government assistance program at the adult age.

The Carolina Abecedarian Project selected a group of 111 children from families in risk situations, with low schooling levels' parents, low salaries and high levels of health problems. Children were selected in the first months of life and randomly separated into two groups. The treatment groups had intensive attention for eight hours per day, five days per week until reaching 5 years. When the children reached 5 years old there was another drawing for the formation of another treatment groups until 8 years old. At the end of the program there were, thus, four groups. One group that received the treatment in the two periods, one group that received the treatment in the first period but not the second, another group that did not receive treatment in the first period but did in the second and a group that did not receive treatment in either period. This division in four groups allowed the evaluation of whether the treatment received in the first period continued to have an effect even if interrupted.

These two programs had in common a small sample design, which does not permit the generalization of their results. A high-level program was Head Start, which began in 1965. The program was aimed at 3 to 5 year old children, guaranteeing access to preschool to disadvantaged children. The results of the evaluations show positive effects on university entrance and lower rates of criminality.

For Latin America, Mexico's Oportunidades, Bolivia's Proyecto Integral de Desarrolo Infantil and the analysis of the construction of preschools in Argentina stand out. For the program Oportunidades, the results indicate an improvement in the motor activities of children, an increase in the probability of preschool enrollment and expected years of education for the treatment group (Gertler and Fernald, 2004; Behrman, Parker and Todd, 2004; Schady, 2006). The expanding supply of preschools in Argentina contributes to the improvement in performance in the Spanish language for the beneficiary children (Berlinsk et al., 2006).

For Brazil, the studies that evaluate effects of interventions in early childhood focus on the analysis of the effect of preschool on the proficiency levels of students in primary and secondary education and also on results in the labor market.

Curi and Menezes-Filho (2009) estimate the effect of preschool on the conclusion of academic cycles, on academic performance measured by proficiency tests and on future salaries. They use data from PPV and from SAEB 2003. These authors used the MQO and Probit estimators and found a positive relationship between preschool and future results, though not significant in all the specifications.

Felício and Vasconcellos (2007) evaluate the effect of preschool on proficiency in mathematics of 4th-grade students from public schools using data from SAEB 2003 and Prova Brasil 2005. Based on a matching analysis for the individual data from SAEB 2003 and schools' fixed effects built from the schools' panel data of SAEB 2003 and Prova Brasil 2005, the authors conclude that preschool has positive effect on proficiency.

Calderini and Souza (2009), using data from Prova Brasil 2005, adopt an instrumental variables approach in order to evaluate the effect of preschool on proficiency in mathematics and Portuguese of 4th-grade students and found positive results in this relation.

Felício, Terra and Zoghbi (2012) analyze the relationship between age of entrance in school and literacy measurements in 2nd-grade children in the municipality of Sertãozinho-SP using data from Provinha Brasil. By means of an analysis by matching estimators they concluded that students who enrolled in school at 5 years of age or less obtained higher scores in relation to children who entered school at 6 or older. Another result presented is noteworthy, significant differences were not found between students who attended one year of Child Education and students who attended for more years.

Almeida and Pazello (2010) used data from the 1982 and 2007 PNAD and work with aggregated data from successive educational generations, relating the proportion of individuals from a certain generation that attended preschool and their future academic performance. The authors conclude that enrolling earlier in school has positive effects on the conclusion of primary education, but there are not significant effects on the conclusion of secondary and higher education.

The chart below shows a summary of the principal results found in the empirical literature, as well as the database used and method applied.

[Insert Chart 1]

3 Database and Variables

The database used is GERES, a longitudinal study performed in order to evaluate the quality and equality of Brazilian primary education. The project was coordinated by six Brazilian higher education institutions: Federal University of Bahia – UFBA, Federal University of Juiz de Fora – UFJF, Pontifical Catholic University of Rio de Janeiro – PUC Rio, Federal University of Minas Gerais – UFMG, University of Campinas – Unicamp and State University of Mato Grosso do Sul – UEMS and was funded by the Ford Foundation and PRONEX – Center of Excellence Program/CNPq.

GERES evaluated three hundred three schools in five Brazilian cities over the course of four years. The cities evaluated were Salvador (BA), Belo Horizonte (MG), Rio de Janeiro (RJ), Campinas (SP) and Campo Grande (MS). A sample of students was followed over the course of these four years and five waves of tests with these students were applied. The students selected were studying, in the first year of the project, the second grade of elementary school or the equivalent grade when the educational system still had an eight-year primary school. At the beginning of the first year a test was applied in order to serve as a baseline and an adjustment for the other tests. After this first wave, the tests were administered at the end of each year. Due to changes of schools it was not possible to follow all of the students initially registered. The city of Salvador only participated in the first four waves.

Despite the longitudinal characteristic of GERES database, some questionnaires were not applied in all years, except the one related to the teacher. Then it is not possible to accomplish neither the evolution of many inquiries about school's characteristics nor the socio-economic level of families. This limitation of database prevents us to monitoring the evolution of investments in schools and family inputs. In order to dealing with such a restriction we sought information about schools in the schools census, conducted annually by Anísio Teixeira National Institute of Research and Studies – INEP. On the other hand, an important variable considered in the analysis of learning is the teacher's salary. This information was obtained from RAIS-MIGRA, matching schools' CNPJ we could get the average salary of teacher per school. For public schools the average salary is the average for county and school system.

The selection of the variables used as inputs in the educational production function was based on several studies on the effect of characteristics of the teachers and of the schools assembled on the Itaú Social Foundation site, in particular in the Factors Associated with Educational Success project. Chart 2 shows the variables used and the source database .

[Insert Chart 2]

4 Empirical Strategy

The estimation of the educational production function considering the theory of child development requires information about the entire history of investments in the child, from the first years until school age, besides a measurement of the innate ability of the child. Due to the limitation of the existing databases not having all of this information, several strategies have been adopted in the literature. Todd and Wolpin (2003) perform a review of the main methods utilized, as well as the need for information and hypotheses of identification strategy for each method to obtain consistent estimations. The section briefly presents the overall model used in Todd and Wolpin (2003) and the empirical strategy used in this paper.

Equation (1) shows the overall form, where Aia represents the skill level of individual i at age a, $F_i(a)$ represents the history of the family's investments until age a, $S_i(a)$ the school inputs until age a, u_{i0} indicates innate ability, acquired at the moment of birth, and f_a is a function that relates the inputs to the formation of skills. Overall, this function can vary with age.

$$A_{ia} = f_a[F_i(a), S_i(a), u_{i0}] \tag{1}$$

To simplify the display of the strategy, school and family inputs are represented by a single vector $X = (F, S)$.

Considering that the technology of skill formation is the same for all ages, $f_a = f$. And considering that f is a linear and additive function separable in its arguments, (1) can be rewritten as (2):

$$A_{ia} = X_{ia}\alpha_1^a + X_{i,a-1}\alpha_2^a + X_{i,a-2}\alpha_3^a + ... + X_{i,1}\alpha_a^a + \beta_a u_{i0} + \varepsilon_i(a)^1 \tag{2}$$

where α represents the effect of X on the ability level, the superscript indicates the effect of the age input a and the subscript indicates the distance between the period in which the input was applied and the period in which proficiency is measured. According to Todd and Wolpin (2003), Equation (2) represents the true model.

The main difficulties for the estimation of (2) are: i) scarcity of databases containing information about the entire history of investment in the child; ii) lack of a measurement of innate ability acquired at the moment of birth; iii) possible correlation between innate ability and the level of investments.

The correlation indicated in item (iii) is verified if we consider the optimizing behavior of families. In this scenario, the choice of investment level in children' skills formation depends on the level of innate ability.

Considering specifically the case of preschool, enrollment in early childhood education is not mandatory in Brazil, this decision is a free choice of the parents. This fact may produce a correlation between non-observed abilities of the child and preschool attendance. Rodrigues *et al.* (2010) indicated four possible channels through which the correlation between the decision to enroll in preschool and the score on proficiency tests can occur; (i) if the families that enroll their children in preschool are families that value value more human capital investment, the children that attend preschool will have greater stimulus to study and thus will obtain higher proficiency levels not due to preschool; (ii) it is possible that parents with higher opportunity costs of time are those who send their children in preschool in order to exploit their advantages in the labor market. Considering that some of these advantages are also important in the formation of the child's human capital (education of the parents), it is possible that part of the difference between those that went to preschool and those that did not is due the fact that parents, in the first case, present higher level of human capital; (iii) if the parents that enroll the children in preschool are those with less ability to care for their children, it is possible that part of the difference between the proficiency levels of the children that attend preschool and those that do not is, in part, from the inability of the parents to promote the formation of children's human capital; and (iv) it is possible that sending children to preschool is a way of compensating deficiencies in child development, which would mean that

[1]The subscript 1 indicates the effect of input applied throughout the year in proficiency measured at the end of the year. The subscript 2 indicates the effect of the previous year and so the following

the children with less ability are those who are enrolled in preschool. If one of these hypotheses occurred, the simple difference between the proficiency levels of the children that attended preschool and those that did not would not reflect the causal effect of preschool. In the first two cases, this difference will be overestimated whereas in the last two it will be underestimated.

Despite the empirical and theoretical evidence of the correlation between the preschool variable and the innate ability of the child, several studies ignore this fact and use strategies adopting the hypothesis of non-correlation. Therefore, the problem of estimation is restricted to the absence of information about the history of investments. Two strategies are indicated in the literature.

The first strategy is to ignore the learning dynamic and to consider that only the contemporaneous inputs determine learning, this strategy is known as contemporaneous specification. And the second strategy, known as value added, is possible to be applied when there is at least one lagged measure of proficiency. This lagged measurement captures the history of investments.

The simplest strategy that can be adopted in terms of the quantity of information is the contemporaneous specification, however, it is the one that demands a greater number of restrictions on the identification hypotheses. This specification considers only the inputs applied during the year in which proficiency is being measured and adopts the following hypotheses: i) only the contemporaneous inputs matter, or the past inputs are not correlated with the present inputs; ii) the contemporaneous inputs are not correlated with non-observed innate ability. The contemporaneous specification assumes the form defined in Equation (3):

$$A_{ia} = X_{ia}\alpha + \varepsilon_i(a) \tag{3}$$

In this specification the error term includes all the past inputs, innate ability and measurement errors. If hypotheses (i) and (ii) in the previous paragraph are satisfied, the estimation by OLS will produce consistent estimates.

Due to the difficulties in observe the entire history of children's human capital investment and the evidence that the hypotheses of the contemporaneous specification seldom hold, many studies now use a strategy known as value added. For this strategy, we need to observe one lagged measurement of proficiency that is assumed to capture the entire past history of investments. As can be seen by comparing Equations (3) and (4), the only difference from the value added to the contemporaneous specification is the presence of lagged proficiency.

$$A_{ia} = \gamma A_{1,a-1} + X_{ia}\alpha + \varepsilon_{ia} \tag{4}$$

Despite this estimation considering that the learning process is a dynamic phenomenon, and, consequently approaching the true model of skill formation, the estimation of this model by OLS requires the non-presence of non-observed abilities, even if the non-observed abilities are not correlated with the level of inputs of vector X. If there is the presence of non-observed abilities in the formation of skills, these skills will be correlated with the lagged proficiency measurement.

Owing to the limitations of the contemporaneous and the value added specifications, Todd and Wolpin (2003) proceed to outline the properties of estimation of the cumulative model. If the choice of inputs is not correlated with non-observed innate ability, the estimation of (2) by OLS produces consistent estimates. However, if the choice of the parents and of the school is correlated with the skill level the estimates will be inconsistent.

To control for the correlation between non-observed ability and the inputs, at least two observations at different periods of time are necessary. An estimate of the fixed effects is used in order to expunge the effect of the non-observed innate abilities. Todd and Wolpin (2003) detail the hypotheses necessary for obtaining consistent estimates using the estimation of the cumulative model with fixed effects.

The GERES database does not have information about the entire history of investments; it only has information about the four first years of the school period and an evaluation at the beginning of the school period, a diagnostic assessment. This assessment is used for supplement the absence of investment information in the period preceding the academic one. Therefore, taking into account the information available in the GERES database and the strategies previously presented, in this article a mixed strategy is adopted, with cumulative specification and value added.

The mean proficiency of the fifth wave is defined as A_{ia}. There is, therefore, vector $A = (A_{ia},\ A_{i,a-1},\ A_{i,a-2},\ A_{i,a-3},\ A_{i,a-4})$ and vector $X = (X_{ia},\ X_{i,a-1},\ X_{i,a-2},\ X_{i,a-3})$. Besides these two information vectors, it is also possible to know if the child has attended preschool. Preschool is an input applied before $X_{i,a-3}$. Therefore, $A_{i,a-4}$ captures the information of vector $X\prime = (X_{i,a-4}, X_{i,a-5}, ..., X_{i1})$ and whether the child has attended or not preschool. Thus, it is assumed that the effect of vector $X\prime$ on proficiency, measured during the school term, is manifested through $A_{i,a-4}$. This is an adaptation of the hypothesis of dynamic completeness. Formally,

$$E\left[A_{ia}|X_{ia}, X_{i,a-1}, ..., X_{i1}, u_{i0}\right] = E\left[A_{ia}|X_{ia}, X_{i,a-1}, X_{i,a-2}, X_{i,a-3}\, , A_{i,a-4}, u_{i0}\right]$$

..

Considering the four years of elementary school in which the students are followed in GERES, the model of skill formation for each year can be visualized by means of the system of equations defined in (5).

$$A_{ia} = X_{ia}\alpha_1^a + X_{i,a-1}\alpha_2^a + X_{i,a-2}\alpha_3^a + X_{i,a-3}\alpha_4^a + \gamma^a A_{i,a-4} + \beta^a u_{i0} + \varepsilon_{ia}$$

$$A_{i,a-1} = X_{i,a-1}\alpha_1^{a-1} + X_{i,a-2}\alpha_2^{a-1} + X_{i,a-3}\alpha_3^{a-1} + \gamma^{a-1} A_{i,a-4} + \beta^{a-1} u_{i0} + \varepsilon_{i,a-1} \qquad (5)$$

$$A_{i,a-2} = X_{i,a-2}\alpha_1^{a-2} + X_{i,a-3}\alpha_2^{a-2} + \gamma^{a-2} A_{i,a-4} + \beta^{a-2} u_{i0} + \varepsilon_{i,a-2}$$

$$A_{i,a-3} = X_{i,a-3}\alpha_1^{a-3} + \gamma^{a-3} A_{i,a-4} + \beta^{a-3} u_{i0} + \varepsilon_{i,a-3}$$

According to what is presented in (5), the preschool input would have its effect only upon the level of proficiency in the diagnostic assessment and therefore it would not be possible to identify it with this strategy. However, studies show that the fact of the child attending preschool increases cognitive and non-cognitive capabilities and these capabilities continue to have effects on learning. For example, the increase in non-cognitive skills such as motivation and/or concentration can make the children's learning in the school term greater than the learning of children who did not attend preschool. The new hypothesis of dynamic completeness can be written as in (6).

$$E\left[A_{ia}|X_{ia}, X_{i,a-1}, ..., X_{i1}, \Pr e_i, u_{i0}\right] = E\left[A_{ia}|X_{ia}, X_{i,a-1}, X_{i,a-2}, X_{i,a-3}, A_{i,a-4}, \Pr e_i, u_{i0}\right] \qquad (6)$$

With this new hypothesis defined, the system of equations defined in (5) comes to be written by the new system defined in (7).

$$A_{ia} = X_{ia}\alpha_1^a + X_{i,a-1}\alpha_2^a + X_{i,a-2}\alpha_3^a + X_{i,a-3}\alpha_4^a + \gamma^a A_{i,a-4} + \delta^a \Pr e_i + \beta^a u_{i0} + \varepsilon_{ia}$$

$$A_{i,a-1} = X_{i,a-1}\alpha_1^{a-1} + X_{i,a-2}\alpha_2^{a-1} + X_{i,a-3}\alpha_3^{a-1} + \gamma^{a-1} A_{i,a-4} + \delta^{a-1} \Pr e_i + \beta^{a-1} u_{i0} + \varepsilon_{i,a-1} \qquad (7)$$

$$A_{i,a-2} = X_{i,a-2}\alpha_1^{a-2} + X_{i,a-3}\alpha_2^{a-2} + \gamma^{a-2}A_{i,a-4} + \delta^{a-2}\Pr e_i + \beta^{a-2}u_{i0} + \varepsilon_{i,a-2}$$

$$A_{i,a-3} = X_{i,a-3}\alpha_1^{a-3} + \gamma^{a-3}A_{i,a-4} + \delta^{a-3}\Pr e_i + \beta^{a-3}u_{i0} + \varepsilon_{i,a-3}$$

With the system of equations defined in (7), it is possible to estimate the effect of the inputs applied during the school term, the effect of initial ability level of the child and the direct effect of preschool on each year in which the child is evaluated. However, the fact of the child having attended preschool can increase the skill level with which the child begins elementary school, in other words, part of the preschool effect is captured by the measurement of proficiency in the diagnostic assessment, an indirect effect. Thus, the effect obtained with this strategy could be underestimating the real effect.

Todd and Wolpin (2003) adopt the hypothesis of constant age effect, that is, $\alpha_x^a = \alpha_x^{a'}$. This assumption reduces the number or parameters. Here this hypothesis is relaxed, allowing the effect of the input to vary with the age. In this context it is possible to verify, for instance, if a "good teacher" has a different impact on the child's learning in the first grade and in the fourth grade. In order to implement such a strategy we introduce variables of interaction between inputs and a wave dummy that indicates the wave in which proficiency is being measured. As an example α_1^a measures the effect of an input applied in the wave a on the child's proficiency observed in the same wave; α_1^{a-1} indicates the effect of the same input applied in the wave $a-1$ on the child's proficiency achieves on wave $a-1$; α_2^a represents the effect of an input applied in the wave $a-1$ on the child's proficiency in wave a. The system of equations in (8) shows the modified system in (7) with the inclusion of interaction variables:

The estimation of the system presented in (8) by OLS will produce inconsistent estimates if the non-observed innate ability is correlated with the other inputs. To control for this possible endogeneity, the strategy of fixed effects is used by an estimator of first difference.

$$A_{ia} = (X_{ia} * W_a)\alpha_1^a + (X_{i,a-1} * W_a)\alpha_2^a + (X_{i,a-2} * W_a)\alpha_3^a + (X_{i,a-3} * W_a)\alpha_4^a + \gamma^a(A_{i,a-4} * W_a) + \delta^a(\Pr e_i * W_a) + \beta^a u_{i0} + \varepsilon_{ia}$$

$$A_{i,a-1} = (X_{i,a-1} * W_{a-1})\alpha_1^{a-1} + (X_{i,a-2} * W_{a-1})\alpha_2^{a-1} + (X_{i,a-3} * W_{a-1})\alpha_3^{a-1} + \gamma^{a-1}(A_{i,a-4} * W_{a-1}) + \delta^{a-1}(\Pr e_i * W_{a-1}) + \beta^{a-1}u_{i0} + \varepsilon_{i,a-1} \quad (8)$$

$$A_{i,a-2} = (X_{i,a-2}*W_{a-2})\alpha_1^{a-2} + (X_{i,a-3}*W_{a-2})\alpha_2^{a-2} + \gamma^{a-2}(A_{i,a-4}*W_{a-2}) + \delta^{a-2}(\Pr e_i*W_{a-2}) + \beta^{a-2}u_{i0} + \varepsilon_{i,a-2}$$

$$A_{i,a-3} = (X_{i,a-3} * W_{a-3})\alpha_1^{a-3} + \gamma^{a-3}(A_{i,a-4} * W_{a-3}) + \delta^{a-3}(\Pr e_i * W_{a-3}) + \beta^{a-3}u_{i0} + \varepsilon_{i,a-3}$$

To purge the effect of observed abilities from the system defined in (8) it is necessary to assume the hypothesis that their effect would not vary with age, in other words, $(\beta^a = \beta^{a-1})$, in this case the differentiation eliminates non-observed ability.

5 Descriptive Statistics

This section presents some descriptive statistics from the GERES data. Table 1 shows the mean and standard deviation of the variables used separately both for the students who attended preschool and for the students who did not.

It can be seen that the students who did preschool have, on average, a higher skill level during the four first years of the academic period. In mathematics the mean proficiency is 192.4 for the students who attended preschool and 166.4 for those who did not, a difference of 26 points on the proficiency scale, or 0.37 standard deviations. For Portuguese, the mean values are 151.8 and 139.4, for those who attended preschool and those who did not, respectively, which corresponds to a difference of 0.42 standard deviations. This difference cannot be interpreted as the effect of preschool in function of non-randomness in the preschool selection.

With respect to the characteristics of the students, it is perceived that approximately 35% of the students were white, but this percentage of whites is less for those who did not attend, 28%, in other words, white children had greater access to preschool. In relation to gender there is certain homogeneity, with a slightly higher attendance rate for girls. Socioeconomic level presents a greater inequality between the two groups (attended and did not attend preschool). Children who attended preschool had a higher socioeconomic level than those who did not, which indicates a positive selection with respect to attending preschool.

The information of the teachers showed that the teachers of students that attended preschool had, in general, higher salaries, more time in the profession and a higher percentage of teachers with higher education. As for the schools, 82% of the students who attended preschool enrolled in public schools and 95% of those who did not attend preschool went to public schools. Scholastic resources, computer labs and libraries are well distributed between the groups; only the presence of a science lab is more significant for the group of students who did preschool, 30%, versus 17% for those who did not. The last line in the table shows that 87% of the children in the sample attended preschool.

In general, the results presented in Table 1 indicate that the students who attended preschool had a higher level of investments during the first four years of primary school, both for academic inputs and for private inputs, considering the socioeconomic levels of the families as a proxy.

[Insert Table 1]

Graphs 3 and 4 show the evolution of proficiency in mathematics and Portuguese, respectively, for the students who attended preschool and for those who did not. The behavior of the two disciplines is different, however, there is a feature in common, in both the group of students that attended preschool starts elementary school with higher proficiency scores than the group of students who did not attend preschool and continues with higher scores during the four years evaluated. The difference is that for mathematics the evolution is relatively parallel, it is only between the 2nd and 3rd wave that the group that did not attend preschool has a lesser evolution, but later the development rate recovers and continues increasing at rates close to the group of students that did preschool. For Portuguese, from the beginning of the series the development rate of the group that attended preschool is greater, making the distance between the groups expand.

[Insert Graph 3 and Graph 4]

Graph 5 shows the relationship between the child's attendance in preschool and the mother's education. The relationship is strictly increasing, that is, the greater the mother's education, the higher the percentage of children enrolled in preschool. Considering the evidence of a positive effect of preschool on learning, this fact implies the generation of a vicious cycle, in which children of parents with a low level of education have a lower probability of achieving academic success.

[Insert Graph 5]

6 Results

This section presents the econometric results of the paper. First, we discuss the results of preschools effects estimates on proficiency in Mathematics. Then we compare these estimates with the results related to proficiency in Portuguese. Our estimates are presented in a sequence. First we report an exploratory estimation with a contemporaneous specification for Ordinary Least Squares. Afterwards, the cumulative estimations are presented with and without considering the value added. These specifications are estimated by the OLS estimator and First Difference.

For each specification four models are utilized with the gradual inclusion of controls. Model 1 includes as controls only characteristics of the child and the family, more precisely race of the student and socioeconomic level of the family. In Model 2 the school system is added, whether it is public or private. In Model 3 characteristics of the infrastructure of the school are used: the presence of a computer lab, a science lab, a library and Internet access. And finally, in Model 4 variables with characteristics of the teachers, experience, salary and education, are added.

Table 2 shows the first results for proficiency in Mathematics. Only the coefficients of the variable of interest, preschool, are reported. This first specification separately shows the effect of preschool in each wave. It can be seen that the effect in wave 4 is greater than in wave 2, indicating that the effect of preschool persists throughout the first years of elementary school and that the distance in proficiency between those who attended preschool and those who did not increases.

Comparing the four models, it can be noted that the magnitude of the preschool coefficient decreases as controls related to the school and the teachers are included. This fact indicates that there is a positive correlation between the preschool variable and the school inputs, in other words, the children who attended preschool form a group that receives better school inputs. For the estimations in this specification to be consistent it is necessary to assume that there is not a correlation between the inputs present and the past inputs, nor between the inputs present and non-observed innate ability.

[Insert Table 2]

Table 3 shows the first results of the cumulative specification. In general, the coefficients present a smaller magnitude in relation to the coefficients shown in Table 4. This shows that the hypothesis necessary for the estimation by contemporaneous specification of non-correlation between inputs present and past inputs is not verified, in other words, the formation of knowledge is a dynamic process, as indicated by the Theory of the Lifecycle of the Skill Formation of Cunha and Heckman (2007).

[Insert Table 3]

Table 4 shows the results for the mixed specification, cumulative with value added. The inclusion of proficiency from the diagnostic assessment has the objective of capturing the effect of the inputs used before the academic period and not observed in the GERES database. In order for this specification to obtain consistent estimates, it is necessary that there is not an effect from non-observed abilities on the skill formation.

The results found present a lower magnitude with respect to those presented in Table 3. This decrease in magnitude indicates a positive correlation between past inputs, not observed, and inputs applied during the academic period.

[Insert Table 4]

To control for possible correlations between the level of investments and innate non-observed ability the next two specifications are done by means of the technique of first differences.

In estimations by First Difference we need to omit one of two interactive dummies between preschool and wave and also between initial proficiency and wave. The interactions with wave 2 were omitted in the two cases. Therefore, the coefficients of the other interactions should be analyzed as the difference in relation to the omitted dummy. In Table 5 the results show that the effect of preschool on proficiency in Mathematics is greater in waves three, four and five than its effect in wave two.

[Insert Table 5]

Table 6 reports the results for the estimation of fixed effect with value added. Focusing the analysis on Model (4), it can be seen that only the coefficient of wave three is significant, that is, the students who attended preschool gained more skills than the students who did not attend preschool until wave three, afterwards the difference between the two groups stabilizes.

Comparing the results from Table 6, cumulative specification with value added and fixed effects, with the results from Table 5, cumulative specification with fixed effect without value added, it can be seen that the inclusion of initial proficiency as an independent variable made the coefficients of the interactive preschool dummies decrease in magnitude in wave two and lose significance in waves three and four. This indicates that it is not possible to accept the hypothesis of non-correlation between the inputs applied in the academic period and the inputs invested before entrance of the child into school, nor the hypothesis of non-correlation between the decision of the parents to enroll their children in preschool and the other investments of the parents in the skill formation of their children before their entrance in school.

The coefficients of the interactive variables between initial proficiency and test administration wave, with positive and significant values, shows that learning development in Mathematics during the academic period depends on the skill level with which the child starts elementary school. The coefficients increase in magnitude over the years. In reinforces the assumptions made by Cunha & Heckman (2007)'s model. In this model learning is a step that depends on the level of learning in the previous step. This implies that an inequality of Mathematical skills in children tends to increase in the academic period.

[Insert Table 6]

The following tables show the results for proficiency in Portuguese. Table 7 shows the results for the estimate by OLS of the contemporaneous specification for the four years of the GERES project. In this estimation, the effect of preschool is positive and significant for all the waves analyzed and for all the models. The same occurred for proficiency in Mathematics, however, the magnitude of the effect is smaller for Portuguese and has similar values in all the waves, in Mathematics the effect increased over the years. In the same manner as for mathematics, the inclusion of controls reduces the effect of preschool, which indicates the correlation between the level of inputs applied during the academic period and the preschool variable.

[Insert Table 7]

Table 8 presents the results for the cumulative specification without value added estimated by OLS. The same comments made about the results in Table 7 are valid. Similar to Mathematics, the results are positive and significant, but the magnitude is less and more stable over the years.

[Insert Table 8]

In order for the results in Table 8 to be consistent, the inputs invested before the entrance of the child in school should be orthogonal to the inputs invested in the academic period. With the estimation using the strategy of value added it is possible to test this hypothesis. If the coefficients of initial proficiency, which capture the information about investments in the period preceding school, are significant, it is not possible to assume the hypothesis of orthogonality between inputs from the academic period and from past inputs. Table 9 shows that the coefficients are positive and significant. Therefore, part of the preschool effect presented in Table 8 is, in fact, an effect from the skills with which the child entered school.

[Insert Table 9]

Table 10 brings the results for estimation controlled for fixed effects, without value added. In all the models specified and for all the waves, the effects were not significant.

[Insert Table 10]

Table 11 presents the results for the estimation of value added controlling for non-observed effects. Again, the results of the preschool dummies were not significant. The results for initial proficiency are noteworthy. Differently from what happened with the results in Table 6 for proficiency in Mathematics, the effect of initial proficiency decreased over the years, even becoming negative in waves 4 and 5 . This result shows that for Portuguese the skill level with which the child starts school does not determine academic success, that is, even if there is an inequality of skills in Portuguese at the moment in which a child starts elementary school, this inequality tends to decrease if the academic inputs are distributed efficiently.

[Insert Table 11]

7 Final Considerations

The aim of the paper was to estimate the effect of preschool attendance on the level of proficiency in Mathematics and Portuguese. To achieve this objective a longitudinal database that followed a sample of children from four Brazilian municipalities over the course of the first four years of elementary school was used. The children evaluated were those who, in 2005, the first year of the project, attended the 1st grade of elementary school.

The effect of preschool was estimated based on an educational production function taking into account the dynamic learning process. The empirical strategy used was an adaptation of estimation by value added, in order to consider the cumulative effect of the inputs and the presence of non-observed heterogeneities. To verify how the hypotheses utilized in the more restrictive strategies, contemporaneous specification and value added, do not hold, the results for all the specifications were presented gradually.

The results found indicate that the most adequate specification is that which considers the cumulative effect of the inputs and controls for individual non-observed heterogeneity. The effect of preschool differs between the proficiencies in Mathematics and Portuguese. For Mathematics the direct effect found was positive and significant. As to the indirect effect, manifested through initial proficiency, the effect found, besides positive, increases over the years.

In the case of Portuguese, the preschool coefficients display non-significant results, indicating that there is not a direct effect on skills in Portuguese. As for the indirect effects, the students with higher skills at the beginning of elementary school expanded their differences in relation to less skilled students, but this effect inverts over the years, indicating that the inequality of skills in Portuguese can be reduced during the schoolar period.

As a subside to public policies for reducing inequalities and increasing the population level of education, the results of this article corroborate the evidence of investments need in early childhood, mainly for the increase of skills in Mathematics. Considering that preschool attendance is higher for children of higher socioeconomic level' families, a policy of child education aimed at poorer families could contribute to decreasing the effect of the vicious cycle of skill formation, reducing the effect of social origin upon children's learning.

References

[1] Almeida, R. B. and Pazello, E. T. (2010). O efeito da pré-escola sobre o desempenho escolar futuro dos indivíduos. *Proceedings of the 38th Brazilian Economics Meeting*, ANPEC - Brazilian Association of Graduate Programs in Economics.

[2] Barnett, W. S. (1992). Benefits of Compensatory Preschool Education. *Journal of Human Resources*, University of Wisconsin Press, vol. 27(2), pages 279-312.

[3] Barnett, W. S. (1995). Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes. *Future of Children*. Volume 5 Number 3 Winter.

[4] Behrman, J., Parker, S. W. and Todd, P. E. (2004). Medium Term Effects of the Oportunidades Program Package, Including Nutrition, on Education of Rural Children Age 0-8 in 1997. *Technical Document. Number 9 on the Evaluation of Oportunidades.*

[5] Berlinski, S., Galiani, S. and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, Elsevier, vol. 92(5-6), pages 1416-1440, June.

[6] Berlinski, S., Galiani, S. and Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, Elsevier, vol. 93(1-2), pages 219-234, February.

[7] Casella, E. B., Amaro Jr, E. and Costa, J. C. (2011). As Bases Neurológicas da Aprendizagem da Leitura. In: *Aprendizagem Infantil: Uma abordagem da neurociência, economia e psicologia cognitiva.* Aloisio Araújo (coord.). Rio de Janeiro.

[8] Blau, D., and Currie, J. (2006). Preschool, daycare, and afterschool care: Who's minding the kids? In E. Hanushek, & F. Welch (Eds.), *Handbook of the Economics of Education*, Handbooks in Economics, vol. 2, chap. 20. Amsterdam: North-Holland, pp. 1163-1278.

[9] Calderini, S. R. and Souza, A. P. (2009). Pré-escola no Brasil: Seu Impacto na Qualidade da Educação Fundamental. *Proceedings of the 37th Brazilian Economics Meeting*, ANPEC - Brazilian Association of Graduate Programs in Economics. Foz do Iguaçu, Paraná, p. 1-21.

[10] Cunha, F. and Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2): 31–47.

[11] Cunha, F. and Heckman, J. (2009) The economics and psychology of inequality and human development. *NBER Working Paper* no. 14695. 2009.

[12] Cunha, F. and Heckman, J. (2010). Investing in Our Young People. *NBER Working Paper* No. w16201.

[13] Cunha, F., Heckman, J. , Lochner, L. J., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek, & F. Welch (Eds.), *Handbook of the Economics of Education*, chap. 12. Amsterdam: North-Holland, pp. 697-812.

[14] Cunha, F., Heckman, J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78 (3), 883-931.

[15] Curi, A. and Menezes-Filho, N. A. (2009). A Relação entre Educação Pré-Primária, Salários, Escolaridade e Proficiência Escolar no Brasil. *Estudos Econômicos*, São Paulo, v. 39, n. 4, p. 811-850.

[16] Currie, J. (2001). Early Childhood Education Programs. *Journal of Economic Perspectives*, American Economic Association, vol. 15(2), pages 213-238, Spring.

[17] Felício, F. de and Vasconcellos, L. (2007) O Efeito Da Educação Infantil Sobre O Desempenho Escolar Medido Em Exames Padronizados. P*roceedings of the 35th Brazilian Economics Meeting*, ANPEC - Brazilian Association of Graduate Programs in Economics.

[18] Felicio, F., Terra, R. and Zoghbi, A. C. (2012). The effects of early childhood education on literacy scores using data from a new Brazilian assessment tool. Estud. Econ., São Paulo, v. 42, n. 1, Mar.

[19] IBGE - Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios – PNAD, 2001-2009. Rio de Janeiro.

[20] GERES, (2005). Estudo Longitudinal sobre a Qualidade e Equidade no Ensino Fundamental Brasileiro.

[21] Gertler, P. and Fernald, L. (2004). The Medium Term Impact of Oportunidades on Child Development in Rural Areas. Unpublished manuscript, University of California at Berkeley.

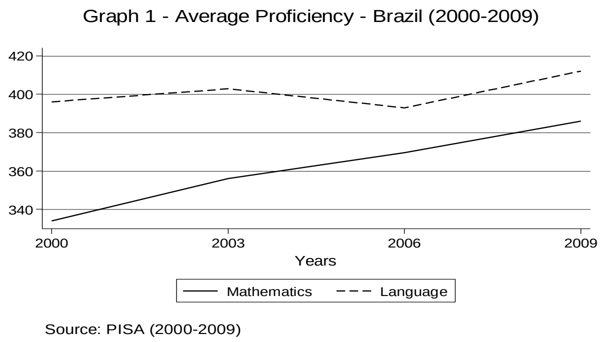
[22] Knudsen, E. I. (2004). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, 16(1), 1412{1425.

[23] PISA – Programme for International Student Assessment. (2009).

[24] Rodrigues, C.G, Pinto, C. X. C. and Santos, D. D. (2010). The impact of daycare attendance on mat test scores for a cohort of 4 graders in brazil. Working paper.

[25] Todd. P. E. and Wolpin, K. I. (2003). On The Specification and Estimation of The Production Function for Cognitive Achievement. *Economic Journal*, Royal Economic Society, vol. 113(485), pages F3-F33, February.

[26] Schady, N. (2006). Early Childhood Development in Latin America and the Caribbean. *Economía* 6(2): 185-213.



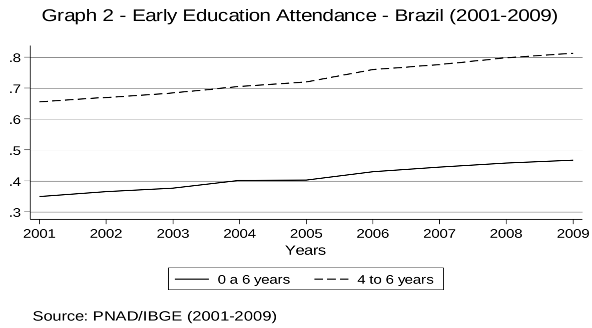


Chart 1 – Review of main empirical brazilian studies of preschool effect

Authors	**Datasets**	**Estimation method**	**Main results**
Curi e Menezes-Filho (2009)	SAEB 2003	Ordinary Least Square	Positive effects, not significant
Felício and Vasconcellos (2007)	SAEB 2003 and Prova Brazil 2005	Matching and school fixed effects	Positive effects
Calderini and Souza (2009)	Prova Brazil 2005	Instrumental variable	Positive effects
Felício *et al* (2009)	Provinha Brazil	Matching	Positive effects
Rodrigues *et al* (2010)	SAEB 2005	Instrumental variable	Positive effects

Source: Authors' Elaboration.

Chart 2 – Variables Used in the Empirical Analysis

Variables	**Data set source/Year**
Mathematics and Language proficiency	Geres/2005-2008
Child and family variables	
Race	Geres/2008
Socioeconomic level	Geres/2008
Teacher variables	
Education e experience	Geres/2005-2008
Teacher wage	Rais-Migra/2005-2008
School variables	
Computer lab, Science lab and Librarie	Censo Escolar/2005-2008
Public school/Private school	Geres/2005-2008

Source: Authors' Elaboration.

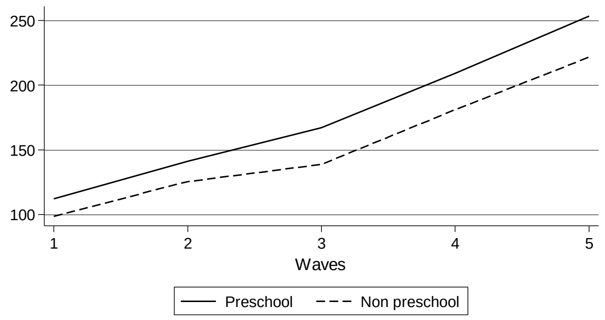
Table 3 – Mean and standard error

	Preschool attendance (A)	Non Preschool attendance (B)	Difference (A – B)
Mathematics Proficiency	152.002	139.460	12.542***
	(29.551)	(28.136)	
Language Proficiency	192.718	166.698	26.020***
	(69.646)	(63.965)	
Socioeconomic level	0.164	-0.216	0.380***
	(0.619)	(0.472)	
White	0.349	0.282	0.067***
	(0.477)	(0.450)	

Race not reported	0.006	0.010	-0.004***
	(0.080)	(0.102)	
Boys	0.492	0.533	-0.042***
	(0.500)	(0.499)	
Girls	0.500	0.455	0.046***
	(0.500)	(0.498)	
Gender not reported	0.008	0.012	-0.004**
	(0.091)	(0.110)	
Public school	0.813	0.948	-0.136***
	(0.390)	(0.221)	
Computer lab	0.723	0.753	-0.031***
	(0.448)	(0.431)	
Science lab	0.303	0.169	0.134***
	(0.460)	(0.375)	
Libraries	0.739	0.721	0.017**
	(0.439)	(0.448)	
Teacher experience - 0 to 5 years	0.060	0.058	0.001
	(0.237)	(0.235)	
Teacher experience - 6 to 10 years	0.116	0.120	-0.004
	(0.321)	(0.325)	
Teacher experience – 11 to 15 years	0.159	0.153	0.006
	(0.366)	(0.360)	
Teacher experience – more than 15 years	0.519	0.468	0.052***
	(0.500)	(0.499)	
Teacher experience – not reported	0.146	0.201	-0.056***
	(0.353)	(0.401)	
Teacher with higher education	0.706	0.665	0.041***
	(0.456)	(0.472)	
Teacher wage	2,160.15	2,065.95	94.197***
	(1,288.73)	(1,027.83)	
Teacher wage – not reported	0.024	0.012	0.011***
	(0.152)	(0.110)	
Belo Horizonte city	0.283	0.302	-0.019**
	(0.451)	(0.459)	
Campinas city	0.301	0.244	0.057***
	(0.459)	(0.430)	
Campo Grande city	0.124	0.297	-0.173***
	(0.330)	(0.457)	
Rio de Janeiro city	0.292	0.156	0.135***
	(0.454)	(0.363)	

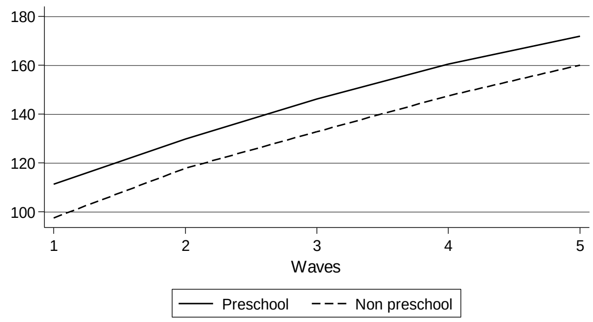
Source: Geres (2005-2008)
Note: Standard errors in parentheses.

Graph 3 - Mathematics Proficiency (2005-2008)



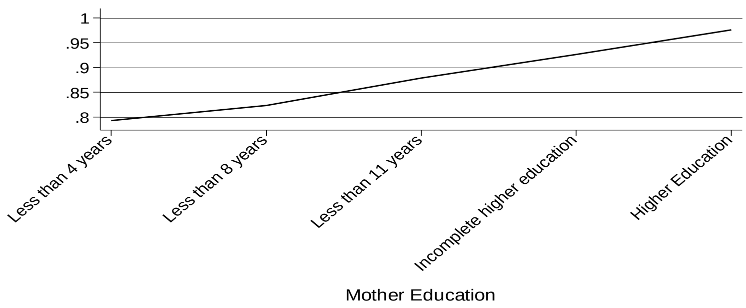
Source: Geres(2005-2008)

Graph 4 - Language Proficiency (2005-2008)



Source: Geres(2005-2008)

Graph 5 - Preschool attendance and mother education



Source: Geres(2005-2008)

Table 4 – OLS – Mathematics

Dependent variable– Mathematics Proficiency	Wave2	Wave3	Wave4	Wave5
		Model 1		
Preschool	6.573***	12.055***	10.650***	13.090***
	(0.979)	(1.615)	(1.799)	(1.956)
Observations	8,476	8,476	8,476	8,476
R-square	0.219	0.248	0.229	0.228
		Model 2		
Preschool	6.472***	11.866***	10.495***	12.940***
	(0.965)	(1.592)	(1.783)	(1.938)
Observations	8,476	8,476	8,472	8,476
R-square	0.240	0.269	0.243	0.242
		Model 3		
Preschool	6.019***	10.803***	9.803***	12.229***
	(0.949)	(1.555)	(1.784)	(1.927)
Observations	8,452	8,381	8,079	8,334
R-square	0.274	0.312	0.272	0.254
		Model 4		
Preschool	5.158***	10.076***	9.088***	9.906***
	(0.941)	(1.544)	(1.772)	(1.888)
Observations	8,452	8,381	8,079	8,334
R-square	0.290	0.323	0.283	0.289

Note 1: Model 1: race and socioeconomic level; Model 2: model1 + public/private school; Model 3: model 2 + computer lab, science lab, library, internet; Model 4: model 3 + teacher wage, education and experience.
Note 2: ***Significant at 1%. **Significant at 5%. *Significant at 10%.
Note 3: Standard errors in parentheses.

Table 5 – Cumulative Specification - OLS without Value Added – Mathematics Proficiency

Dependent variable– Mathematics Proficiency	(1)	(2)	(3)	(4)
Preschool*wave2	6.573***	6.472***	5.992***	5.131***
	(1.630)	(1.613)	(1.577)	(1.539)
Preschool*wave3	12.055***	11.881***	10.518***	9.498***
	(1.630)	(1.613)	(1.578)	(1.542)
Preschool*wave4	10.650***	10.495***	9.245***	8.285***
	(1.630)	(1.613)	(1.580)	(1.544)
Preschool*wave5	13.090***	12.944***	11.299***	7.811***
	(1.630)	(1.613)	(1.580)	(1.548)
Observations	33,904	33,904	33,904	33,904
R-square	0.513	0.523	0.549	0.572

Notes: see notes to table 4

Table 6 – Cumulative Specification - OLS with Value Added – Mathematics Proficiency

Dependent variable– Mathematics Proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	2.641**	2.661**	2.378*	2.148
	(1.342)	(1.338)	(1.324)	(1.310)
Preschool*wave3	6.066***	6.097***	5.258***	4.828***
	(1.342)	(1.338)	(1.325)	(1.313)
Preschool*wave4	4.258***	4.276***	3.517***	3.289**
	(1.342)	(1.338)	(1.327)	(1.315)
Preschool*wave5	6.578***	6.624***	5.247***	3.413***
	(1.342)	(1.338)	(1.327)	(1.317)
Initial proficiency*wave2	0.753***	0.741***	0.726***	0.718***
	(0.017)	(0.018)	(0.018)	(0.018)
Initial proficiency*wave3	1.147***	1.124***	1.086***	1.069***
	(0.017)	(0.018)	(0.018)	(0.018)
Initial proficiency*wave4	1.225***	1.208***	1.166***	1.141***
	(0.017)	(0.018)	(0.018)	(0.018)
Initial proficiency*wave5	1.248***	1.227***	1.219***	1.140***
	(0.017)	(0.018)	(0.018)	(0.018)
Observations	33,904	33,904	33,904	33,904
R-square	0.671	0.673	0.683	0.691

Notes: see notes to table 4

Table 7 – Cumulative Specification – First difference without Value Added – Mathematics Proficiency

Dependent variable– Mathematics Proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	omitted			
Preschool*wave3	5.482***	5.399***	4.742***	4.296***
	(1.163)	(1.162)	(1.161)	(1.161)
Preschool*wave4	4.078**	4.011**	3.356**	3.130*
	(1.645)	(1.643)	(1.642)	(1.643)
Preschool*wave5	6.517***	6.426***	5.351***	4.392**
	(2.015)	(2.012)	(2.012)	(2.014)
Observations	25,428	25,428	25,428	25,428
R-square	0.553	0.554	0.560	0.563

Notes: see notes to table 4

Table 8 – Cumulative Specification – First difference with Value Added – Mathematics Proficiency

Dependent variable– Mathematics Proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	omitted			
Preschool*wave3	3.425***	3.427***	2.982***	2.748**
	(1.150)	(1.149)	(1.150)	(1.150)
Preschool*wave4	1.617	1.606	1.189	1.171
	(1.626)	(1.625)	(1.627)	(1.628)
Preschool*wave5	3.936**	3.924**	2.938	2.271
	(1.992)	(1.991)	(1.993)	(1.996)
Initial proficiency*wave2	omitted			
Initial proficiency*wave3	0.394***	0.383***	0.363***	0.354***
	(0.015)	(0.015)	(0.015)	(0.016)
Initial proficiency*wave4	0.472***	0.467***	0.446***	0.448***
	(0.021)	(0.021)	(0.022)	(0.022)
Initial proficiency*wave5	0.494***	0.486***	0.496***	0.490***
	(0.026)	(0.026)	(0.027)	(0.028)
Observations	25,428	25,428	25,428	25,428
R-square	0.565	0.566	0.570	0.573

Notes: see notes to table 4

Table 9 – OLS – Language

Dependent variable– language proficiency	**Wave2**	**Wave3**	**Wave4**	**Wave5**
	Model 1			
Preschool	5.201***	5.753***	5.898***	5.206***
	(0.654)	(0.786)	(0.781)	(0.746)
Observations	8,476	8,476	8,476	8,476
R-square	0.259	0.234	0.210	0.202
	Model 2			
Preschool	5.123***	5.670***	5.831***	5.158***
	(0.642)	(0.777)	(0.774)	(0.741)
Observations	8,476	8,476	8,472	8,476
R-square	0.286	0.251	0.224	0.213
	Model 3			
Preschool	4.853***	5.063***	5.479***	4.947***
	(0.627)	(0.761)	(0.781)	(0.738)
Observations	8,452	8,381	8,079	8,334
R-square	0.327	0.292	0.244	0.225
	Model 4			
Preschool	4.312***	4.765***	5.179***	3.910***
	(0.623)	(0.757)	(0.776)	(0.716)
Observations	8,452	8,381	8,079	8,334
R-square	0.340	0.300	0.256	0.275

Notes: see notes to table 4

Table 10 – Cumulative Specification - OLS without Value Added – Language Proficiency

Dependent variable– language proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	5.201***	5.123***	4.875***	4.330***
	(0.744)	(0.736)	(0.720)	(0.705)
Preschool*wave3	5.753***	5.676***	5.033***	4.745***
	(0.744)	(0.736)	(0.720)	(0.706)
Preschool*wave4	5.898***	5.827***	5.497***	5.048***
	(0.744)	(0.736)	(0.721)	(0.707)
Preschool*wave5	5.206***	5.165***	4.697***	3.379***
	(0.744)	(0.736)	(0.721)	(0.709)
Observations	33,904	33,904	33,904	33,904
R-square	0.444	0.456	0.484	0.507

Notes: see notes to table 4

Table 11 – Cumulative Specification - OLS with Value Added – Language Proficiency

Dependent variable– language proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	1.501**	1.537**	1.573***	1.378**
	(0.597)	(0.597)	(0.596)	(0.590)
Preschool*wave3	1.802***	1.824***	1.601***	1.458**
	(0.597)	(0.597)	(0.596)	(0.591)
Preschool*wave4	2.363***	2.378***	2.384***	2.048***
	(0.597)	(0.597)	(0.597)	(0.592)
Preschool*wave5	2.043***	2.062***	1.829***	1.069*
	(0.597)	(0.597)	(0.597)	(0.593)
Initial proficiency*wave2	0.606***	0.596***	0.582***	0.576***
	(0.009)	(0.009)	(0.009)	(0.009)
Initial proficiency*wave3	0.647***	0.640***	0.624***	0.616***
	(0.009)	(0.009)	(0.009)	(0.009)
Initial proficiency*wave4	0.579***	0.573***	0.562***	0.556***
	(0.009)	(0.009)	(0.009)	(0.009)
Initial proficiency*wave5	0.518***	0.515***	0.513***	0.481***
	(0.009)	(0.009)	(0.009)	(0.010)
Observations	33,904	33,904	33,904	33,904
R-square	0.644	0.645	0.649	0.657

Notes: see notes to table 4

Table 12 – Cumulative Specification – First difference without Value Added – Language Proficiency

Dependent variable– language proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	omitted			
	0.552	0.542	0.272	0.294
Preschool*wave3	(0.500)	(0.500)	(0.502)	(0.502)
	0.697	0.692	0.684	0.622
Preschool*wave4	(0.708)	(0.707)	(0.710)	(0.711)
	0.004	0.015	-0.149	-0.251
Preschool*wave5	(0.867)	(0.866)	(0.869)	(0.872)
Observations	25,428	25,428	25,428	25,428
R-square	0.476	0.477	0.480	0.482

Notes: see notes to table 4

Table 13 – Cumulative Specification – First difference with Value Added – Language Proficiency

Dependent variable – language proficiency	**(1)**	**(2)**	**(3)**	**(4)**
Preschool*wave2	omitted			
Preschool*wave3	0.301	0.276	0.055	0.083
	(0.500)	(0.500)	(0.502)	(0.503)
Preschool*wave4	0.862	0.830	0.809	0.698
	(0.708)	(0.708)	(0.710)	(0.712)
Preschool*wave5	0.542	0.503	0.270	0.093
	(0.867)	(0.867)	(0.870)	(0.872)
Initial proficiency*wave2	omitted			
Initial proficiency*wave3	0.041***	0.044***	0.039***	0.040***
	(0.007)	(0.007)	(0.008)	(0.008)
Initial proficiency*wave4	-0.027***	-0.023**	-0.022**	-0.014
	(0.010)	(0.010)	(0.011)	(0.011)
Initial proficiency*wave5	-0.088***	-0.081***	-0.075***	-0.069***
	(0.012)	(0.013)	(0.013)	(0.014)
Observations	25,428	25,428	25,428	25,428
R-square	0.480	0.481	0.483	0.485

Notes: see notes to table 4