PRE-SCHOOL CONTRIBUTIONS TO FUTURE ACHIEVEMENTS

Luis Fernando Gamboa

SERIE DOCUMENTOS DE TRABAJO

No. 156

Abril de 2014

Pre-school contributions to future achievements.*

Luis Fernando Gamboa†
Facultad de Economía
Universidad del Rosario
Bogotá, Colombia

May 8, 2014

Abstract

This paper provides recent evidence about the benefits of attending preschool on future performance. A non-parametric matching procedure is used over two outcomes: math and verbal scores at a national mandatory test (Saber11) in Colombia. It is found that students who had the chance of attending preschool obtain higher scores in math (6.7%) and verbal (5.4%) than those who did not. A considerable fraction of these gaps comes from the upper quintiles of student's performance, suggesting that preschool matters when is done at high quality institutions. When we include the number of years at the preschool, the gap rises up to 12% in verbal and 17% in math.

JEL: C14, I21, O54

Keywords: Preschool, Education, Colombia

1 Introduction

Investments on human capital during early childhood provide high returns given that cognitive problems are harder to solve as the pupils grow up (Heckman 2000, 2008). Recent efforts have been focused on increasing human capital in developing countries, but these efforts have been primarily devoted to improve the availability of inputs that only tend to help students who already are high achievers.

The number of years of education has been central in the explanation of income distributions, but today some concerns are still latent. On the one hand, the main concern comes from the importance attributed to "*additional diplomas*". It is clear that specialization acquired during masters or Ph.D programs provides important and necessary skills for several occupations but it is also true that the competencies achieved during the first ages are crucial for future performance. On the other hand, not all the students would be able to complete their basic education in developing countries. This occurs mainly because households face income restrictions, absence of basic capabilities from early education (read and math), the lack of schools and preschools in remote areas and teacher's teaching practices.

*The author gratefully acknowledges financial support from Universidad del Rosario and the outstanding research assistance provided by Erika Londoño. The access to database provided by the ICFES was crucial for this document.

†Corresponding author. Address: Calle 12C No. 4-69. Phone 571-2970200. ext 120 Bogotá-Colombia. E-mail: luis.gamboa@urosario.edu.co

This document aims to study the importance of having studied in a pre-school on some outcomes as achievement in test scores (an indicator of future performance in the labor market) and the duration of educational cycle. The score obtained in a national compulsory test is used as an outcome that reflects quality of education. We are interested on identifying whether preschools explain heterogeneity on test scores after controlling by potential confounding factors.

The methodology followed through this document assesses whether preschool explains average differences in the test by means of a non-parametric method that allows us to decompose the differences between observed and unobserved factors. This methodology has been proposed by Ñopo (2008) and applied to different topics (gender gaps, motherhood gaps, regional gaps among others). Although, there are multiple approaches to evaluate gaps using decomposition methods, we opt for using Ñopo since the size of the data does not allow us to have enough variability as it is requested on quantile regression or other procedures (see Fortin *et al.,* 2011).

We find that there is a score gap of 6.7% in math and 5.4% in verbal between students who attended preschool and those who did not. When we measure an intensity effect by including the number of years at the preschool, this gap rises up to 12% in verbal and 17% in math.

The structure of the document is described as follows. Section 2 summarizes the recent literature about the effects of preschool on academic achievements or other educational goals. Section 3 describes the methodology. Section 4 presents the data and the results of the empirical strategy. Section 5 briefly concludes.

2 The Preschool and its implications

During early years, parents and preschools provide basic skills that are crucial for future performances but in many cases these sources of skills are not always available. Following the *Education for All -Global Monitoring Report* from UNESCO (2011), early childhood care and education services help build skills when children's brains are developing. However, there is a wide gap in enrollment between the richest and poorest as a consequence of the cost of private provision, which is one of the factors that contribute to inequity in access at this level, and the availability of educational programs.

Preschool education has expanded considerably. The global pre-primary education gross enrollment ratio increased from 33% in 1999 to 50% in 2011 according to UNESCO(2011). However, this expansion seems to be insuficient. High fixed cost prevents investments in preschools to be undertaken in small cities. It is also possible that in some cases parents prefer to keep their children at home until the age of entry to first grade. The problem arrises from the implications of this choice. Early interventions in human capital include the provision of care centers for stimulating children, the design of programs of improve learning and health practices toward parents and the creation of preschool institutions to develop cognitive and non cognitive skills. This document will be focused on this last type of intervention (see Peisner-Feinberg *et al.*, 2001, for a discussion about it). Burchinal *et al.* (1997) suggest that those who attend a preschool before mandatory basic primary accumulate more tools for developing their skills on math and verbal assignments. This hypothesis comes from the fact that the structure of brain is formed during the first years and there are several ways to foster individual capabilities more efficiently during this age such as self-regulation, motivation, reactivity to stress and discipline. As Hazarika and Viren (2013) suggest, early life experiences strongly affect brain development. There is enough evidence about the advantages and skills obtained from preschool education. Most of the research deal with the

importance of the formation of cognitive and noncognitive skills during the first years. (Currie, 2001; Cunha *et al.*, 2006; Glewwe and Jacoby, 1995).

Recently, providing early education at home as a consequence of differences with educational systems implies a trade off between having the children at school or at home. For example, the benefits of staying at home depend on the quality of the time shared with the parents or other adults. In Hazarinka and Viren (2013) words, children unexposed to language during a sensitive period of early childhood may become incapable of using language appropriately. The acquisition of language and math may also influence the formation of other cognitive and non cognitive skills. Preschool education can help those parents who are not capable of providing good skills for their children. Early treatments at home can also help when specific things are needed or when parents are committed to teaching strategies and practices.

The provision of preschool education in Colombia is carried out in two alternative ways. On the one hand, each school includes up to three years before the first grade. On the other hand, there are small educative institutions that only offer preschool education and send their students to mandatory education in other schools. However, in small and remote cities it is probably the case that there simply are no preschools.

This indicates the importance of having evidence about the benefits of education in children for the case of Colombia. Colombia has faced several economic and political problems limiting the participation of an important fraction of the population into the educational system. Some of them are caused by absence of schools and some others are caused by income restrictions. During the last decade, educational policy has encouraged the retention of students in the schools by reducing demand and supply barriers. Mandatory education in Colombia is composed by a basic cycle (five primary years and four secondary years) and a middle cycle (two years). In the publicly provided education, it is not mandatory to have a specific number of education as it is demanded in private schools.[1] Using data from the National Statistics Office (DANE), preschool enrollment grew from 430.739 students in 1991 to 1.060.269 in 2010. This means a annual growth rate of 4.85% during two decades, but most of this growth took place in the first decade since during 2000-2010, preschool enrollment was very stable in most of the cities. This growth means that while there was about 7 students in primary per each one in preschool in 1991, this ratio is closest to 4 in 2010.

The literature on this field is diverse. Some works try to provide insights about the benefits of receiving specific treatments or interventions during the first years (Barnett and Escobar, 1987; Berlinski *et al.*, 2008, Nores and Barnett, 2010). Other studies focused on the estimation of impact of preschool and its duration on outcomes such as social and cognitive development, verbal and math proficiency and labor productivity, among other factors (Heckman and Masterov, 2004; Barnett and Lamy, 2006; Loeb *et al.*, 2007; Magnuson *et al.*, 2007a; Temple and Reynolds, 2007). Reynolds (1995) investigated the effects of the length of preschool (a Head Start-type program) for several cognitive and social outcomes with respect to other kindergartens that acts as control groups. His findings suggests that while 2-year participants began and finished kindergarten more academically competent than 1-year participants, through the elementary grades these children did not significantly or meaningfully differ from one another in reading comprehension and mathematics achievement, rates of grade retention and special education placement. The overall effect size for grades 1 to 6 was 0.15 standard deviations and values consistently favored the 2-year group. Both 1- and 2-year preschool participants were consistently and significantly better adjusted than no-preschool participants.

[1] Most private schools restrict enrollment to those to attend to more than two years in their own school or at partner intitutions. Some schools have an additional year at the end of middle education.

Caught *et al* (1994) mention that children belonging to appropriate care programs during their first years obtained higher scores in reading comprehension. Recent literature seeks to identify short-term effects of different lengths of exposure to preschool. Some of them suggest that children with longer exposure to preschool demonstrate advantages (e.g. stronger cognitive skills, improved socio-emotional outcomes) over children with shorter preschool exposure, (Loeb et al., 2007; Behrman *et al.*, 2004; Skibbe *et al.*, 2011; Barnett and Lamy, 2006). Nores and Barnett (2010) carry out a meta analysis on the effects of preschool and their results indicate that programs lasting one to three years had average effect sizes of 0.3 standard deviations, as compared to 0.2 for programs lasting less than one year. The returns of an early education investment are seen not only in the short term, but also in the long run. Using the same data Connolly *et al.* (1991), Robertson and Symons (1996) and Harmon and Walker (1998) found that results of children in cognitive test are positively related with their earnings when they are working. In contrast, some authors such as Magnuson et al. (2007a, 2007b) state that the advantages of early education and care will decrease by the second or third grade of formal school. Even so, there is enough evidence in favor of positive but moderate effects of preschool on the cognitive development of the children in the long run (Anderson *et al.,* 2003; Barnett, 1995, 2008; Currie, 2001; Nelson *et al.*, 2003).

However, studies on the link between early enrollment and future grade attainment are not frequent in Latin American countries. The case of Colombia is particularly interesting as a result of its economic and political problems faced during the last decades. These features have prioritized the available budget to war instead of other dimensions as health and education. Bernal and Camacho (2012) mention that low attendance to formal education institutions is common around the country and varies according to socioeconomic status. People from the bottom of the pyramid do not attend or go to low quality schools while children from high income parents attend better quality schools that include various preschool years. Recent initiatives such as "*Familias en Acción*" program foster children participation and permanency at school system.

The absence of programs designed for low income households motivated the governmental social project called "*Hogares Comunitarios*". This program is mainly targeted to vulnerable children with low incomes, in areas such as education, nutrition and other things that could affect academic performance. Its impact is assessed by Bernal and Camacho (2010) who find that, depending on the discount rate used, the cost-benefit analysis shows that there's always a positive return which could be even 8 times the invested amount. This modality of public interventions allow the low income households to increase their opportunities.

Equality of opportunities exists where everyone has the same chance to develop her capacities regardless their circumstances (race, religion, gender, skin color and so on). Following this idea, the childhood with socioeconomic disadvantages are more likely to repeat grades or drop out from the school before finishing their educational cycle.

Thus, early investments on education are important for the governments, when social mobility is a main step to increase the opportunities starting in the early years. Then, public provision of preschool education would help to increase human capital as well as to reduce future inequalities at the labor markets.

3 Methodology

It is not straightforward to identify the effect of one intervention as being part of a preschool when there are multiple unobserved factors. In this specific case, participants were not randomly assigned to each group (control or treatment). The treatment and comparison groups may be similar on

a variety of characteristics but it is not easy to avoid some bias on estimations when there is no experimental design. Controlling for observed variables and reducing the importance of unobserved factors significantly improve the analysis.

The estimation of the importance of doing preschool on test achievements in subjects such as math and verbal is carried out by following the Ñopo (2008) non-parametric procedure. One important challenge of this purpose is to take into account the existence of differences in observable as well as unobserved factors. The choice of math an verbal subjects is done based on the intuition that boys and girls perform differently in these areas of knowledge. This framework starts from comparing individuals with similar characteristics using a matching process based on a five-step algorithm. The first step consists on selecting one student (with no replacement) who attended a preschool (P) from the sample. In the second step, all the students without preschool that share the same characteristics of the student selected in the previous step are selected.

Then, a synthetic individual using all individuals selected in the second step is constructed, with a score (math or verbal) equal to the average score of the selected students. This allows us to match the synthetic student without preschool *(NP)* with the original individual from step 1. The fourth step arranges the observations of the synthetic student and the original student into their new samples of matched student (with and without Preschool). The last step consists on repeating the four steps until the whole original students' sample is exhausted.[2]

Using Ñopo (2008) framework, we have two types of students which are those who do not carried out preschool (No-Prek) and those who studied at preschool (Pre-k). For simplicity, *NP* represents students with no preschool and P are those students that already studied preschool, whose socioeconomic characteristics are denoted by x. Equations (1) and (2) stand for the expected score (in math as well as verbal) of No-Prek and Pre-k students, respectively,

$$E[Y|NP] = \int_{S^{NP}} g^{NP}(x) dF^{NP}(x) \tag{1}$$

$$E[Y|P] = \int_{S^{P}} g^{P}(x) dF^{P}(x) \tag{2}$$

where F^P and F^{NP} are the cumulative distribution functions of student's characteristics, conditional on having been enrolled into a Preschool or not, $S^P$and S^{NP} correspond to the support of the distribution of their characteristics respectively. The gap is defined by $\triangle = E[Y|NP] - E[Y|P]$. Given that $S^P$and S^{NP} are different, each integral of equations (1) and (2) is divided in two parts, within the the common support ($S^{NP} \cap S^P$) and out of the common support ($S^P \cap S^{\overline{NP}}$, $S^{\overline{P}} \cap S^{NP}$)

$$\begin{aligned} \Delta \quad = \quad & \left[\int_{S^{\overline{P}} \cap S^{NP}} g^{NP}(x) dF^{NP}(x) + \int_{S^{NP} \cap S^{P}} g^{NP}(x) dF^{NP}(x) \right] \\ & - \left[\int_{S^{NP} \cap S^{P}} g^{P}(x) dF^{P}(x) + \int_{S^{P} \cap \overline{S^{NP}}} g^{P}(x) dF^{P}(x) \right] \end{aligned} \tag{3}$$

After some algebraic manipulation and redefinition of the integrals in equation (3), the gap is expressed as:[3]

[2] This procedure differs from the propensity score matching (PSM) in that Ñopo (2008) is done using characteristics of individuals instead of propensity scores from logistic regressions.

[3] See Ñopo (2003, 2008) to check the whole procedure.

$$\Delta = \Delta_{NP} + \Delta_P + \Delta_x + \Delta_0 \tag{4}$$

where Δ_{NP} is the part of the gap explained by the differences in characteristics between unmatched and matched student without preschool. This is the weighted difference between the expected outcome (score) of NP out of the common support minus the expected scores of NP in the common support.

$$\Delta_{NP} = \left[\int_{\overline{S^P}} g^{NP}(x) \frac{dF^{NP}(x)}{\mu^{NP}(\overline{S^P})} - \int_{S^P} g^{NP}(x) \frac{dF^{NP}(x)}{\mu^{NP}(S^P)} \right] \mu^{NP}(\overline{S^P}) \tag{5}$$

The weight, $\mu^{NP}(\overline{S^P})$, is the probability measure of the set $\overline{S^P}$ under the distribution $dF^{NP}(.)$ of No-PreK's characteristics, or the probability measure (under the distribution of No-PreK's characteristics) of the sets of characteristics that Pre-K's do not reach. Analogous interpretations could be done for $\mu^P(S^{NP})$. We have that $\mu^{NP}(\overline{S^P}) + \mu^P(S^{NP}) = 1$.

The second term, Δ_P, is the part of the gap (weighted difference between the expected scores) that can be explained by the differences in characteristics between students that already take preschool courses (matched and unmatched).

$$\Delta_P = \left[\int_{S^N} g^P(x) \frac{dF^P(x)}{\mu^P(S^{NP})} - \int_{\overline{S^{NP}}} g^P(x) \frac{dF^P(x)}{\mu^P(\overline{S^P})} \right] \mu^P(\overline{S^{NP}}) \tag{6}$$

Δ_P would be zero either if all students with PreK can be matched to those who do not have PreK. Note that Δ_{NP} and Δ_P are not symmetrically defined. It means that they can not be read as complements each other. Even though this is just a result of the algebraic manipulation found in Ñopo (2003), we suggest that the symmetry can be seen in a different way: we expect that individuals whose scores are represented by the first terms of equations (5) and (6) are the most advantaged of their type.

The term Δ_x is the portion which can be explained by differences in the distribution of characteristics of both types of students on the common support.

$$\Delta_X = \int_{S^{NP} \cap S^P} g^{NP}(x) \left[\frac{dF^{NP}}{\mu^N(S^P)} - \frac{dF^P}{\mu^P(S^{NP})} \right](x)$$

As for Δ_0, it is the unexplained part of the 'score' gap, the component not originated by differences in characteristics of students. If there exists some type of genetic differences or unobserved behavior over the students, Δ_0 would capture this issue, as well as the existence of unobserved characteristics penalized or rewarded by the educational system. However, this method does not allow us to differentiate what means each proportion of the unexplained part of the gap [4].

This approach reduces the biases caused by unobserved heterogeneity, and provides us more accurate measures of the gaps. As we mentioned before, our main focus is on the gaps resulting from differences in math and verbal achievement at SABER 11 test. This test can be used as a measure of long run effects in education since it is presented at least 11 years after finishing preschool.

[4] For details see Ñopo (2003). In brief, $\Delta_O = \int_{S^{NP} \cap S^P} \left[g^{NP}(x) - g^P(x) \right] \frac{dF^P(x)}{\mu^P(S^{NP})}$

Data

Colombia has a mandatory educational system based on five years of primary and six years of secondary (three of basic and three of middle education). During these eleven years, students are required to approve a series of subjects. However, there is no information about what happens during the cycle.[5] The competencies acquired during mandatory education is assessed at the end of the last year in the nationwide mandatoy test known as Saber 11. The ICFES (*Instituto Colombiano para la Evaluación de la Educación*) designs and manages this test with questions about subjects as math, reading and verbal analysis, natural sciences and other optative areas. This test is commonly used by private universities to assess potential freshmen and allocate scholarships. Although there is no evidence of the opposite, it is possible to assume that students perform as good as they can in this test, since obtaining a good score could increase the chance of entry to higher education.

In addition to test scores, the ICFES also collects information about socioeconomic conditions, physical capital and parents schooling. During 2013 year, an additional module of questions about previous conditions with information on the number of schools where the pupil studied, the number of years at a preschool and the year in which some courses were studied (first, fifth and sixth grades) were added to the questionnaire. This set of questions was asked to a random sample from Saber 11 takers in 2013. The empirical exercises are done by exploiting a random sample of students who take this test. Saber 11 database includes $516,471$ observations for 2013 (March and August) and the number of people who answer the subsample of questions related to preschool incidence, -which is equivalent to the 10 percent of students-, was about $59,357$ students.

However some of these observations present problems that limit their use into the empirical exercises. Missing information on was the first criteria for excluding it from the database. Then, we exclude those students for which the reported information present codification problems. For example, students who report ages above 80 or doing basic and middle education in less than three years do not allow us to have accurate control about its true conditions. The final size of the sample includes $49,033$ students after this depuration process. In order to provide an idea about the randomness of the sample Table 1 presents some descriptive statistics es for the sample and the total database. It is easy to see that in demographic as well as in performance related variables, both samples are equivalent.

The length of the educational cycle is one of the benefits of attending preschool. Figures 1 to 3 show that the distribution of students who attend to a preschool is more concentrated and it also seems to indicate less years on each cycle, on average, than those who does not attend. Additionally, there are gross differences in test scores among students with and without preschool (Figure 4). That is, both distributions are different in math and verbal.

Table 2 and 3 summarize the distribution of some variables used in the matching process and some variables related to educational outcomes for each level of preschool intensity. Our data let us to discriminate students according to the number of years at preschool from 0 to 3 years. Some interesting facts emerge from this table. First, it seems that people with more preschool years tend to spend less years on average in each part of the educational cycle with the exception of primary education. As a result, pupils with no preschool tend to be older when finish secondary education than others. Second, it is unsurprising that students with studies before primary come predominantly from private schools. The supply of public kindergartens is still very low at a national level. Third, performance in math as well as in verbal analysis is higher as preschool

[5] Additionally to SABER 11, the ICFES also carry out a couple of test (Saber 5 y Saber 9) for a sample of students, but it is not possible to follow the evolution of each student in that tests.

increases. There are more than 5 pp between those who had attended in a preschool and those who did not. Finally, parent's schooling is linked to higher performance, on average. This is a common factor in the literature given that more education more attention to children's future. The distribution of the population in each one of the categories allows us to compare the students performance and some of their characteristics (Table 3).

The use of this database presents some drawbacks that are important to mention. First of all, it is not possible to know whether students attend the same institution for preschool and in the basic and middle education. This implies that we only deal with quantity instead quality effect. That is, we do not have detailed information for controlling all the factors that affect the student's performance through the educative cycle, then we only compare number of years at the school. Some researchers mention that preschool provides elements for improving achievements when the student is in a good institution. We do not have information over the educational system that serves to control for short run effects. That is, one student can be enrolled into a high quality preschool and then he moves to a middle quality educational institution for basic and middle education. At the end of the cycle, it is not possible to affirm whether her achievement is explained by more years at preschool or for being few years in a better preschool. Second, we have a selection problem, because we do not have information about those who have already studied a preschool and drop out the educational system before the end of mandatory cycle. Despite of this fact, the probability of completing the eleven years of mandatory education is above the 90% according to data from the "*Gran Encuesta Integrada de Hogares, GEIH*". In consequence, our results have to be read as conditional on completing high school education.

Results

First of all, an ordinary least squares estimation is carried out using scores (math and verbal) as dependent variables (Table 4). Attending preschool is included as a dummy variable and other controls are included. The estimated coefficients suggest that even the inclusion of additional controls, attending a preschool increases the score at SABER 11. Although the estimations for math and verbal are highly similar, it is not clear that they are not biased because of the existence of unobserved factors. Thus, the matching procedure could reduce the size of the bias.

The results of the gap decomposition are summarized in the Table 5 and 6. Table 5 summarizes the results for the entire distribution and Table 6 does it for quantiles of students from the same sample.

There are two sets of results based on the subject used as outcome. The upper side of the Table 5 includes Math as outcome while Verbal is in the panel b. First of all, our results seem to suggest that differences in math are bigger than in verbal. The average gap on math (verbal) when we compare students with and without preK is about 7% (5.4%). We also estimate the score gap between those who carried out 3 years of preschool against people with no preschool. As it is expected, differences increased. The increase is of more than 12% in verbal (17% in math). The size of this gap is considerably high because it reflects that those pupils that did not attended preschool perform 1/6th below those with attend to it. From a complementary point of view, an ordinary least squares estimation of performance is done and the results are summarized in Table7. These results show that the coefficient of dummy variables related to study two or three years of preK are positive significant. There is no significant effect from having one year at preschool. This fact, indicates that the importance of early years of education emerges after the first year.

The following step consist on adding controls to have more comparable students into the matching process. These controls are essentially related to family and school characteristics. The set of controls used for doing the matching process are gender, age, mother schooling, dummy for public school, dummy for calendar of the school, a dummy for those who mention that carried out basic and middle education in the same school and a categorical variable of population size. Each row includes the controls used in the previous one and one additional control. That is, when we control for mother schooling in the case of math, we obtain that 1.4 percentual points remains unexplained from 6.7% of total gap. As a result of this iterative process, when using the complete set of variables the unexplained fraction of the gap is about 0.7 percentual points. This finding implies that there is not an important fraction that we can attribute to unobserved factors using this sample. One additional aspect to highlight is the size of the standard deviation of the unexplained part of the gap. As it can be seen from Table 5, this fraction is statistically significant.

The comparison between people with three years of preschool and students with no preschool presents higher differences, but at the same time exhibits a higher proportion of unexplained gap. This is not an unexpected result because the type of population and the size of the samples (students with 3 years are $8,356$ in the database). These findings are in line with those of Ritblatt (2010)

D_P is the part of the score gap that can be explained by the differences in the characteristics between two groups of the students with preschool, that are in and out of the common support and D_{NP} is the fraction of the average gap explained by differences in characteristics between students without preschool education that are in and out of the common support. As Ñopo (2008) indicates, D_X is the part of the gap that can be explained by differences in the distribution of characteristics of the two groups discriminated in the common support of the variables used into the matching process. The *one-by-one* inclusion of additional controls in math as well as verbal analysis modified the importance of the unexplained fraction of the gap. The left part of the table shows that Dx is the component with higher increases as the number of controls does. The importance of Dx comes from the existence of differences on the distributions between students with and without preschool. The existence of a segmented educational market in which some people are enrolled in distinct schools than others, could help us to explain that preschool is not always available for everyone. As we mentioned before, attending preschool can still be a privilege for a small fraction of students.

In order to provide a more detailed description of this gap, the methodology is also used for assessing whether the gaps are similar or not through the entire distribution. In that case, we split the sample in quintiles according to their performance in math as well as in verbal analysis. Results are summarized in Table 6. Both subjects exhibit similar trends according to the quintile. That is, most of the total gap found in Table 5 comes from the quantile in which high performers are located. The total gap varies strongly from 0.004 in the first quantile to 0.016 in the last quantile of verbal (four times) and from 0.002 to 0.044 in the case of math (more than 10 times). This finding is similar to Bernal and Camacho's findings. That is, preschool helps to increase quality of education when it is provided by high quality institutions. In Colombia, most of these institutions are private kindergartens or private schools that offer more than 14 years of education.

Conclusions

The main challenges of educational policy are those related to the best ways to allocate private and public spending in order to reach better results in terms of equity, quality and efficiency. This document provides new evidence about the importance of first years of education on future achievements. We use a non parametric procedure for decomposing score gaps in academic test

between pupils that attended preschool and those who did not. As it is found in previous literature for other countries, the results obtained in this document suggest that attending preschool can be a good condition for future achievements. These achievements are measured about eleven or twelve years later, then many unobserved factors might affect the final performance, but there is no reason to think that they are not randomly distributed.

The choice of sending the children to the preschool is influenced by the cost-benefit analysis of providing care at home or at the preschool and their availability in some cities. However, many of the choices related to education are in the hands of parents who are not the immediate beneficiaries of these choices. Final achievements are a consequence of previous parental choices. One of them is the type of school that their siblings attend. This choice for instance, is restricted by the household's budget constraint. Today it is very common that private schools offer more years of education before the mandatory cycle as a consequence of their own teaching strategies and their competition for the demand of students.

Our results suggest that average differences can be above 5% of scores obtained at SABER 11 and they are bigger in math than in verbal. When it is controlled by the length of preschool education, this gap increases up to 17% in mathematics.

Most of these differences are in the upper tail of the distribution, which is directly related to the purchase capacity of households located at the top of income distribution, showing that good education requires good schools and preschools. This finding is not small since most of all the private universities in Colombia base their admissions on this test. Moreover, the results obtained in this test could be highly correlated with the admission tests administered by public universities. It is also important to mention that the unexplained part of the score gap tends to be short when the complete set of controls are used. Obviously, this set should be greater but the available information and the size of the sample prevent us from using it. More controls imply small samples in doing the match.

In a country that stivess for better education and more opportunities it is necessary to have better databases to monitor evolution of students achievements through the educational cycle. In addition, initiatives that encourage the creation of guidelines for minimum competencies in preschool programs are needed since the supply of public education before mandatory cycle is only a recent issue. It is necessary to expand the analysis of the early education's short and long effects after the implementation of recent programs such as "De Cero a Siempre" over outcomes such as scores at *Saber 5*, *Saber 9*, the *drop-out* rates and the duration of basic studies.

References

[1] Anderson, L. M., Shinn, C., Fullilove, M. T., Scrimshaw, S.C., Fielding, J. E., Normand J. (2003). The effectiveness of early childhood development programs. American Journal of Preventive Medicine, 24, 32-46.

[2] Barnett, S. W. and Lamy C. E. (2006). Estimated impacts of number of years of preschool attendance on vocabulary, literacy and math skills at kindergarten entry. New Brunswick: National Institute for Early Education Research.

[3] Barnett, S. W. and Escobar C.M. (1987). The economics of early intervention: A review, Review of Educational Research, 57, 387-414.

[4] Barnett, S. W. (1995). Long-term effects of the early childhood programs on cognitive and school outcomes. The future of Children, 5(3), 25-50.

[5] Barnett, S. W. (2008). Preschool education and its lasting effects: Research and policy implications (EPRU Policy Brief). Boulder and Tempe: Education and the Public Interest Center & Education and Policy Research Unit.

[6] Berhman, J. R., Cheng Y. , and Todd P. E. (2004). Evaluating preschool programs when length of exposure to the program varies: a nonparametric approach, The Review of Economics and Statistics 86 (1): 108-132.

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

[8] Bernal, R., and Camacho, A. (2010). La importancia de los programas para la primera infancia en Colombia. Documentos CEDE.

[9] Bernal R., and Camacho A.. (2012) La política de Primera Infancia en el contexto de la Equidad y Movilidad Social en Colombia. October. Universidad de los Andes, Documentos de Trabajo CEDE, p.78.

[10] Burchinal, M. R., Campbell, F. A., Bryant, D. B., Wasik, B. H., & Ramey, C. T. (1997). Early intervention and mediating processes in cognitive performance of children of low-income African-American families. Child Development, 68(5),935–954.

[11] Caught, M . O., Di Pietro, J. A., and Strobino, D . M. (1994) . Day-Care participation as a protective factor in the cognitive development of low-income children. Child Development, 65, 457-471.

[12] Connolly. J, Coopan M., and Kvalsig JD. (1991) The effects of parasite infections on cognitive processes in children. Ann Trop Med Parasitol; 73: 501-6.

[13] Cunha, F., Heckman, J. J., Lochner, L. J., & 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. Amsterdam: North-Holland.

[14] Currie, J. (2001). Early childhood intervention programs: what do we know? Journal of Economic Perspectives, 15, 213-238.

[15] Fortin, N., T. Lemieux and S. Firpo (2011). Decomposition Methods in Economics. In Handbook of Labor Economics. Vol 4a. Ch.1 pp 2-102.

[16] Glewwe, P. and Jacoby H. (1995). An economic analysis of delayed primary school enrollment in a low income country: The role of early childhood nutrition. The Review of Economics and Statistics, 77(1), 156-169

[17] Hazarika G. and Viren V. (2013) The effect of early childhood developmental program attendance on future school enrollment in rural North India. Economics of Education Review 34 146–161

[18] Harmon, C. and Walker I. , (1998). "Selective Schooling and the Returns to the Quantity and Quality of Education: Evidence for men in England and Wales" manuscript Department of Economics, University College Dublin.

[19] Heckman, J. (2000). Policies to foster human capital. Research in Economics, 54, 3–56.

[20] Heckman, J. (2008). Schools Skills and Synapses. NBER Working Papers 14064.

[21] Heckman, J., & Masterov, D. M. (2004). The productivity argument for investing in young children. Working paper 5, Invest in kids working group, Center for Economic Development.

[22] Hazarika, G., and Bedi, A. S. (2009). Child work and schooling costs in rural northern India. In R. Kanbur & J. Svejnar (Eds.), Labor markets and economic development. Oxon/New York: Routledge.

[23] Loeb, S., Bridges M., Bassok D., Fuller B. and Rumberger R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development, Economics of Education Review, 26 (1), 52-66.

[24] Magnuson, K A., Ruhm, C, and Waldfogel, J (2007a) Does prekindergarten improve school preparation and performance? Economics of Education Review, 26, 33-51.

[25] Magnuson, K A., Ruhm, C, and Waldfogel, J. (2007b). The persistence of preschool effects: Do subsequent classroom experience matter? Early Childhood Research Quarterly, 22(1), 18-38.

[26] Nelson, G., Westhues, A., and MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. Prevention and Treatment, 6, 1-35.

[27] Nores, M., and Barnett, S. W. (2010). Benefits of early childhood interventions across the world: (Under)investing in the very young. Economics of Education Review, 29(2), 271–282.

[28] Ñopo, H. (2008). "Matching as a Tool to Decompose Wage Gaps." The Review of Economics and Statistics, 90(2), pp. 290-299.

[29] Ñopo, H. (2003). "Matching as a Tool to Decompose Wage Gaps," Middlebury College Working Paper Series 0406, Middlebury College, Department of Economics.

[30] Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L., Howes, C., Kagan, S.L., and Yazegian, N. (2001). The relation of preschool child-care quality to children's cognitive and social development trajectories through second grade. Child Development, 72 (5), 1534-1553.

[31] Reynolds, A. J. (1995). One year of preschool intervention or two: Does it matter? Early Childhood Research Quarterly, 10(1), 1-31.

[32] Robertson, D. and Symons, J. (1996). "Do Peer Groups Matter? Peer Group Versus Schooling Effects on Academic Attainment," Centre for Economic Performance, London School of Economics, WP No. 311.

[33] Skibbe, L. E., Connor C., Morrison F., and Jewekes A. M. (2011). Schooling effects on preschoolers' self-regulation, early literacy, and language growth, Early Childhood Research Quarterly, 26, 42-49.

[34] Ritblatt, S. N., Brassert, S. ,Johnson R. and Gomez F. (2001). Are two better than one? The impact of years in Head Start on child outcomes, family environment, and reading at home, Early Childhood Research Quarterly, 16: 525-537.

[35] Temple, J. A., and Reynolds, A. J. (2007). Benefits and costs of investments in preschool education: Evidence from the Child-Parent Centers and relat- ed programs. Economics of Education Review, 26(1), 126–144.

[36] UNESCO (2011). The Education for All Global Monitoring Report. Avaliable at http://unesdoc.unesco.org/images/0022/002256/225660e.pdf on April 10th 2014.

Figure 1: Lenght of Educational Cycle

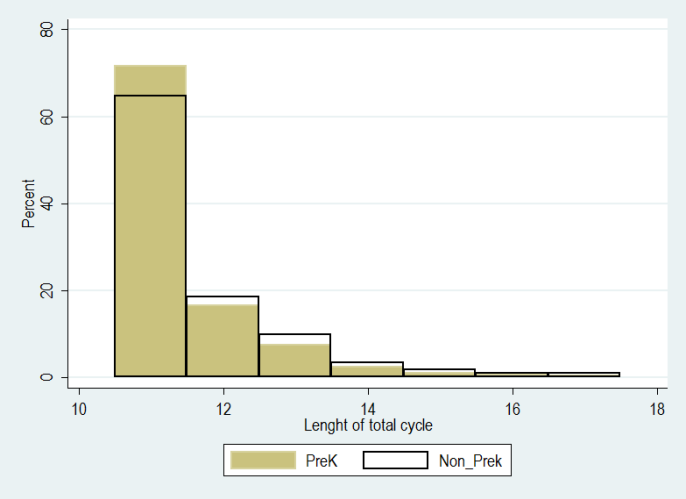

Figure 2: Lenght of Secondary

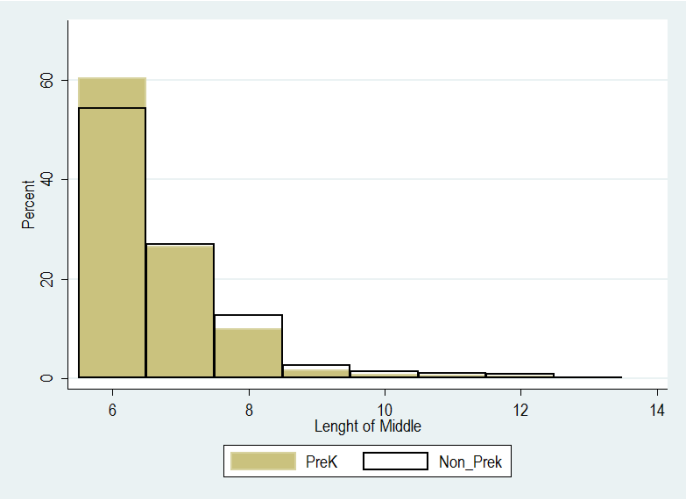

Figure 3: Lenght of Basic Primary (5 years)

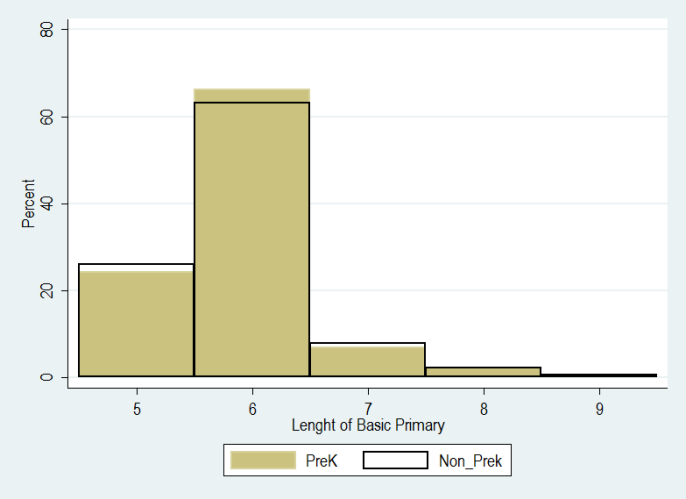

Figure 4: Performance in test scores by intensity of Prek

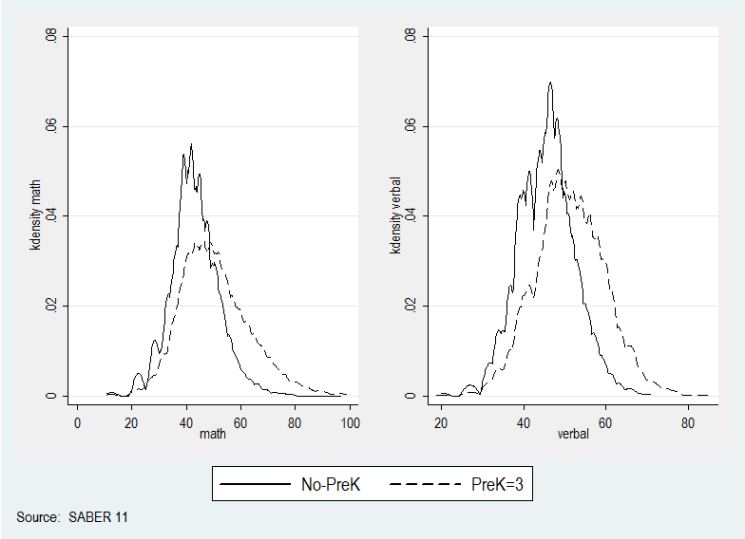

Table 1: Summary Statistics for sample vs Database

	Total	**Sample**
Age	16,98	16,77
Girls	0,54	0,54
Public School	0,72	0,77
Father schooling	4,66	4,69
Mother schooling	4,67	4,72
Math	45,16	45,68
Sd Verbal	7,70	7,71
Verbal	47,38	47,78
Sd Math	10,49	10,70
Observations	495665	49033

Table 2: Summary Statistics by number of years at preK

	PreK=0	**PreK=1**	**PreK=2**	**PreK=3**
Duration years	11,87	11,56	11,50	11,41
Duration Secondary	7,03	6,68	6,61	6,53
Duration Primary	5,89	5,89	5,90	5,88
Age	17,38	16,77	16,60	16,50
Girls	0,53	0,55	0,53	0,54
Public School	0,90	0,88	0,66	0,41
Father schooling	3,63	4,21	5,26	6,42
Mother schooling	3,55	4,25	5,34	6,43
Math	43,08	44,24	47,34	50,54
Sd Math	9,00	9,48	11,45	12,77
Verbal	45,56	46,85	48,92	51,17
Sd Verbal	6,92	7,10	8,08	8,51
Observations	5.609	27.058	8.010	8.356

Table 3: Distribution and averages (PreK vs. No-PreK)

	No-Prek			Prek		
	Math	**Read**	**Obs**	**Math**	**Read**	**Obs**
Gender						
Girls	41,64	45,33	53,1%	44,58	48,00	54,4%
Boys	44,70	45,82	46,9%	47,74	48,14	45,6%
Age						
15	46,62	48,63	7,6%	47,81	49,59	8,9%
16	44,48	46,90	30,8%	47,33	49,24	45,2%
17	42,91	45,41	27,2%	45,51	47,59	27,7%
18	42,07	44,25	17,1%	43,62	46,01	11,7%
19	40,30	43,39	17,4%	40,95	43,43	6,5%
Nature						
Public	42,93	45,44	90,4%	44,27	46,85	74,7%
Private	44,50	46,70	9,6%	51,21	51,66	25,3%
City						
Small	42,39	44,89	56,0%	43,56	46,00	35,4%
Medium	43,65	46,10	25,1%	46,51	48,43	30,7%
Big	44,33	46,82	18,9%	48,14	49,88	34,0%
Mother						
Basic	42,08	44,87	59,4%	42,64	45,37	17,3%
Secondary Incomplete	43,67	45,76	16,3%	44,17	46,94	27,4%
Secondary complete	44,37	46,63	17,7%	45,82	48,06	9,5%
Technical	45,89	48,14	3,3%	49,08	50,79	11,8%
College	49,52	49,98	2,9%	53,24	53,09	2,7%
Postgraduate	47,50	51,45	0,4%	57,71	55,83	2,9%
Father						
Basic	42,47	45,02	65,7%	42,92	45,68	37,9%
Secondary Incomplete	43,28	45,98	13,1%	44,69	47,22	15,3%
Secondary complete	43,93	46,26	14,9%	46,01	48,29	24,9%
Technical	46,49	48,46	2,7%	49,40	50,68	7,2%
College	48,53	49,39	3,3%	52,85	52,82	11,9%
Postgraduate	51,50	53,44	0,3%	59,17	56,60	2,8%
CalendarA	43,05	45,55	99,6%	45,66	47,83	97,2%
CalendarOther	50,09	49,23	0,4%	58,42	56,00	2,8%
Total obs.			5.609			43.424

Table 4: **Determinants of Performance (OLS estimates)**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Math			Verbal		
Prek	0,341***	0,265**	0,260**	0,489***	0,434***	0,431***
	(0,130)	(0,130)	(0,130)	(0,100)	(0,101)	(0,100)
Age	-1,333***	-1,301***	-1,120***	-1,143***	-1,119***	-0,996***
	(0,0429)	(0,0429)	(0,0471)	(0,0317)	(0,0316)	(0,035)
Male	3,199***	3,196***	3,216***	0,234***	0,232***	0,246***
	(0,0901)	(0,0898)	(0,0898)	(0,0648)	(0,0646)	(0,065)
Mother Schooling	0,924***	0,708***	0,707***	0,686***	0,532***	0,531***
	(0,0231)	(0,0261)	(0,0260)	(0,0164)	(0,0185)	(0,019)
Public school	-3,067***	-2,811***	-2,805***	-1,941***	-1,758***	-1,753***
	(0,131)	(0,131)	(0,131)	(0,0910)	(0,0914)	(0,091)
Calendar-A	-7,349***	-7,067***	-7,150***	-4,332***	-4,130***	-4,186***
	(0,426)	(0,423)	(0,422)	(0,295)	(0,293)	(0,292)
City Size	2,60e-07***	2,50e-07***	2,48e-07***	2,85e-07***	2,78e-07***	0,000***
	(1,69e-08)	(1.69e-08)	(1,69e-08)	(1,18e-08)	(1,18e-08)	(0,000)
Father schooling		0,355***	0,353***		0,254***	0,253***
		(0,0218)	(0,0218)		(0,0155)	(0,015)
Years at school			-0,474***			-0,321***
			(0,0468)			(0,036)
Constant	52,19***	51,07***	56,12***	52,27***	51,47***	54,89***
	(0,480)	(0,483)	(0,694)	(0,342)	(0,345)	(0,512)
Observations	48,515	48,513	48,513	48,515	48,513	48,513
R-squared	0,173	0,178	0,179	0,165	0,170	0,171

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

Table 5: Distribution and averages (PreK vs. No-PreK)

	Prek vs No-PreK					Total Prek vs No-Prek				
a. Math	6,7%					17,1%				
controls	D0	Dp	DNP	DX	s.dev D0	D0	Dp	DNP	DX	s.dev D0
male	0,068			-0,001	0,003	0,171			0,000	0,004
+ age	0,049			0,018	0,004	0,172			-0,001	0,003
+ mother schooling	0,014	0,001		0,052	0,004	0,166	0,005	0,000	0,003	0,003
+ public	0,010	0,003		0,054	0,004	0,147	0,015	0,000	0,018	0,003
+ citysize	0,007	0,011	0,000	0,049	0,004	0,136	0,039	0,001	0,015	0,002
+ calendarA	0,007	0,013	0,000	0,047	0,004	0,130	0,052	0,000	0,014	0,002
+ one school	0,006	0,019	0,000	0,042	0,004	0,119	0,077	0,007	0,013	0,002
b. Verbal	5,37%					12,17%				
controls	D0	Dp	DNP	DX	s.dev D0	D0	Dp	DNP	DX	s.dev D0
male	0,054			0,000	0,002	0,122			0,000	0,003
+ age	0,038			0,015	0,003	0,123			-0,001	0,002
+ mother schooling	0,013	0,001		0,040	0,003	0,119	0,002	0,000	0,002	0,002
+ public	0,010	0,002		0,042	0,003	0,107	0,007	0,000	0,011	0,002
+ citysize	0,006	0,007	0,000	0,041	0,003	0,100	0,024	0,002	0,009	0,002
+ calendarA	0,006	0,008	0,000	0,039	0,003	0,097	0,031	0,002	0,008	0,002
+ one school	0,007	0,012	0,000	0,035	0,002	0,089	0,050	0,003	0,007	0,002

Notes: The decomposition of score gaps has been obtained using Ñopo (2008). Each line adds one additional control.

Table 6: Gaps by Quintiles (Pre K vs. No PreK)

quintile	Verbal					math				
	D	D0	Dp	DNP	DX	D	D0	Dp	DNP	DX
1	0,004	0,001	0,000	0,000	0,003	0,002	-0,003	0,003	-0,000	0,003
2	0,002	0,002	0,003	-0,000	0,001	0,001	-0,000	0,000	-0,000	0,001
3	0,002	0,000	0,000	0,000	0,001	0,000	-0,000	0,000	0,000	0,000
4	0,002	0,001	0,000	0,000	0,001	0,003	0,001	0,001	0,000	0,001
5	0,016	-0,002	0,009	0,001	0,008	0,044	0,006	0,019	-0,001	0,019

Notes: The decomposition of score gaps has been obtained using gender, age, the nature of school, city size, parets schooling and the calendar of the school."D" represents the Total Gap at each quintile.

Table 7: Determinants of Performance using PreK intensity

VARIABLES	math		verbal	
	(1)	(2)	(3)	(4)
prek=1	-0,0523	-0,0912	0,262**	0,234**
	(0,132)	(0,132)	(0,102)	(0,102)
prek=2	0,955***	0,839***	0,819***	0,735***
	(0,170)	(0,170)	(0,127)	(0,127)
prek=3	2,088***	1,862***	1,510***	1,347***
	(0,186)	(0,186)	(0,136)	(0,137)
Age	-1,066***	-1,042***	-0,939***	-0,921***
	(0,0370)	(0,0370)	(0,0274)	(0,0273)
Male	3,162***	3,160***	0,201***	0,200***
	(0,0899)	(0,0896)	(0,0647)	(0,0645)
Mother schooling	0,858***	0,664***	0,644***	0,503***
	(0,0234)	(0,0262)	(0,0167)	(0,0187)
Public School	-2,540***	-2,353***	-1,612***	-1,476***
	(0,134)	(0,134)	(0,0946)	(0,0947)
CalendarA	-7,020***	-6,775***	-4,173***	-3,996***
	(0,420)	(0,418)	(0,293)	(0,292)
City Size	2,43e-07***	2,35e-07***	2,70e-07***	2,65e-07***
	(1,70e-08)	(1,69e-08)	(1,18e-08)	(1,18e-08)
One school	0,191*	0,206**	-0,352***	-0,341***
	(0,0986)	(0,0984)	(0,0714)	(0,0712)
Father schooling		0,329***		0,239***
		(0,0218)		(0,0156)
Constant	65,77***	64,44***	64,91***	63,94***
	(0,809)	(0,812)	(0,592)	(0,594)
Observations	48.513	48.513	48.513	48.513
R-squared	0,176	0,180	0,167	0,171

Robust standard errors in parentheses
*** p<0,01, ** p<0,05, * p<0,1

Notes: Variable Prek=1 means one year of preschool education, Prek=2 means two years of preschool and so on.