WPS7574

Policy Research Working Paper 7574

# Access to Pre-Primary Education and Progression in Primary School

## Evidence from Rural Guatemala

*Paulo Bastos*
*Nicolas L. Bottan*
*Julian Cristia*

WORLD BANK GROUP
Development Research Group
Trade and International Integration Team
February 2016

Policy Research Working Paper 7574

## Abstract

Evidence on the impacts of a large-scale expansion in pre-primary education is limited and mostly circumscribed to high- and middle-income nations. This study estimates the effects of such an expansion on progression in primary school in rural communities in Guatemala, where the number of pre-primary schools increased from about 5,300 to 11,500 between 1998 and 2005. Combining administrative and population census data in a difference-in-differences framework, the analysis finds that access to pre-primary education increased by 2.4 percentage points the proportion of students that progress adequately and attend sixth grade by age 12. These positive although limited effects suggest the need for complementary actions to produce substantial improvements in adequate progression.

This paper is a product of the Trade and International Integration Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at pbastos@worldbank.org.

# Access to Pre-Primary Education and Progression in Primary School: Evidence from Rural Guatemala*

Paulo Bastos† Nicolas L. Bottan‡ Julian Cristia§

*Keywords*: Early childhood development, pre-primary education, rural areas.

*JEL classification*: I21, I28.

*We thank the co-editor and an anonymous referee for very helpful comments and suggestions. We are also grateful to Samuel Berlinski and seminar participants at the Research Department of the Inter-American Development Bank, Mind the Gap: From Evidence to Policy Impact, and LACEA-Medellin for very helpful comments.

†Development Research Group, The World Bank, United States. E-mail: pbastos@worldbank.org

‡Department of Economics, University of Illinois at Urbana-Champaign, United States. E-mail: bottan2@illinois.edu

§Corresponding author. Research Department, Inter-American Development Bank, United States. E-mail: jcristia@iadb.org

# 1 Introduction

Many low and middle-income countries are in the process of universalizing pre-primary education. Besides the potential positive impacts on child development and maternal labor supply, the provision of public pre-primary education is attractive because it can be implemented by expanding public primary schools *downwards*, which requires limited infrastructure investments and implies that it can be easily accommodated within existing government structures. Therefore, investments in early childhood development have often been directed to expanding access to pre-primary education (UNESCO, 2006; Engle et al., 2011).[1]

Yet empirical evidence on the effects of a large-scale expansion in pre-primary education is limited and mostly circumscribed to high and middle-income countries. Exploiting the state-by-state expansion of kindergarten in the United States, Cascio (2009) finds that whites affected by the expansion are less likely to drop out from high school and to be incarcerated later in life, but finds no effects for blacks. Berlinski et al. (2008) exploit comparisons across siblings in Uruguay and find that pre-primary attendance significantly increases the probability of attending school by age 15. Focusing on a large pre-primary construction program in Argentina, Berlinski et al. (2009) find that attending a pre-primary induces sizable increases in third-grade test scores and causes improvements in child behavior in class.

There are theoretical reasons to recommend caution in extrapolating these findings to poorer countries. Shifting a child from the home environment to pre-primary education may have a smaller opportunity cost when maternal education is low. On the other hand, the quality of pre-primary education may be considerably lower in less developed countries. The effects of pre-primary enrollment on child development may potentially be negative because they are highly dependent on the quality of both the center attended and maternal time (Baker et al., 2008; Almond and Currie, 2011). Yet it is for policymakers in poorer countries that empirical evidence on the effects of expanding access to pre-primary education is especially warranted: in 2010, average gross enrollment in pre-primary education was 15% in low-income countries, compared with 52% and 82% in middle- and high-income countries, respectively (World Bank, 2012).

We add to this literature by examining the medium term impacts of a large-scale expansion of pre-primary education in rural Guatemala, where large segments of the indigenous population live in poverty.[2] Following the 1996 Peace Accords–that put an end to

[1]Engle et al. (2011) provide a recent and comprehensive review of the literature on early childhood development programs in developing countries.

[2]Recent studies have examined the short-term effects of pre-primary attendance in rural China. Luo et al. (2012) document a positive correlation between pre-primary attendance and educational readiness using cross-sectional data. In turn, Wong et al. (2013) do not find evidence that increasing pre-primary

36 years of civil strife–, the government embarked on an ambitious construction program that increased the number of pre-primary schools from about 5,300 to 11,500 between 1998 and 2005.[3] We combine longitudinal school-level administrative data from 1995-2008 with Population Census data at the community level to estimate the impact of opening a pre-primary in a rural community on primary school enrollment and progression by the age of 12. Precise geographic information makes it possible to construct a panel of rural communities that had one primary school during the period of analysis, but no pre-primary school when the construction program began.

Our identification strategy exploits variation over time in the construction of pre-primaries across communities and the fact that some cohorts therein had access to a pre-primary while others did not. By focusing on rural communities with only one primary school and examining various cohorts over time, we circumvent the potential problem that pre-primary access may affect the composition of students attending primary education. The final selection of beneficiary communities was decided by the central government following an *ad hoc* procedure with no strict guidelines. Still, we apply trimming and propensity score re-weighting in order to obtain a comparison group with similar pre-intervention characteristics to the beneficiary communities.

The research design we adopt is closely related to Berlinski et al. (2009), though it differs in a number of ways. First, we exploit variation in access to pre-primaries using a finer level of aggregation. That is, while Berlinski et al. (2009) estimate effects using variation in access to pre-primary education at the municipality level, we estimate effects using variation at the community level. Hence, we count with sharper variation in the variable of interest (a community either has a pre-primary or not). Second, Berlinski et al. (2009) examine effects on test scores of children in third grade of primary school (who are about 8 years old). In contrast, we assess effects on adequate primary school progression for children that are 12 years old. It is well documented that effects on test scores of early childhood interventions tend to fade out during primary education (Almond and Currie, 2011). Therefore, focusing on progression could be a good proxy to gauge whether investments early in life would produce lasting effects on human capital and productivity. Finally, we focus on a substantially poorer context (rural Guatemala as opposed to Argentina), which may be more relevant to low-income countries.[4]

We find a positive effect on adequate progression (expected grade given age) of about 2.4 percentage points (from a mean of around 12.3% for the beneficiary group). We

attendance affects school readiness in a randomized control trial involving a sample of 150 children.

[3]Chamarbagwala and Moran (2011) provide evidence of a strong negative impact of the civil war on the education of the two most disadvantaged groups, namely rural Mayan males and females.

[4]Our empirical strategy is also inspired in Duflo (2001), who examines schooling and labor market consequences of school construction in Indonesia.

observe that communities that gained access to pre-primary education and those that did not have similar pre-intervention trends. Reassuringly, the results remain robust when controlling for differential trends at various geographic aggregation levels and introducing time-varying attributes. The estimated impacts on adequate progression are found to increase in the average pre-primary take-up rate of the community. Although these effects are important, they suggest that complementary actions need to be undertaken in order to generate sizable improvements in adequate progression.

We proceed by exploring if and how pre-primary access had heterogeneous impacts across subgroups of the population. We find larger effects for communities with higher levels of schooling among adults. This finding is not fully explained by higher take-up rates since these rates do not differ between communities with low and high average education. There are several possible explanations behind this difference in impacts. For example, teacher quality may be higher in high-education communities, or there may exist a complementarity between pre-primary attendance and parental education. We do not find differential effects of pre-primary attendance by gender, percentage of indigenous population or chronic malnutrition.

The paper proceeds as follows. Section 2 provides background before Section 3 describes the data employed in the empirical analysis. Section 4 outlines the research design, Section 5 presents the main results, and Section 6 examines their robustness. Section 7 concludes the paper.

## 2 Background

With an income per capita of $2,623 in 2009, Guatemala is considered to be a low middle-income nation. About half of its 14 million inhabitants live in rural areas. A similar proportion is indigenous. There is significant inequality in the country in a context of high poverty and extreme poverty rates (51% and 15%, respectively), concentrated in the indigenous rural populations. Life expectancy at birth reached 70 years in 2009, compared to 73.3 years in the other Central American nations. The under-5 mortality rate is significantly higher than in the rest of Latin America and the Caribbean (39.8 and 22.5 deaths per 1,000 live births, respectively).

Illiteracy is high: 23.7 percent of individuals aged 18 to 49 have no formal education (Living Standards Measurement Survey, 2006). Primary school coverage is virtually universal, though the quality of education is low. Alvarez and Schiefelbein (2007) find that teachers use inadequate instruction methods given the cultural barriers and sociodemographic context. Primary school repetition rates are high, especially during the initial grades (for example, they reach 30% in first grade). As a result, Guatemala has one of the lowest average accumulated years of education in Latin America (Calderón and Urquiola, 2006).

This poor performance in education appears to be rooted in several factors. Almost 40% of children are not native Spanish speakers and tend to have poor academic performance relative to native-Spanish speakers. The long-lasting civil war had strong negative impacts on the education of the two most disadvantaged groups, namely rural Mayan males and females (Chamarbagwala and Moran, 2011). In rural areas, parental education is low and anecdotal evidence suggests that education is not greatly valued−a child's aspiration is to work in agriculture-related activities or housework (Rodríguez, 2001). Low levels of income seem to play a role when deciding whether to enroll or withdraw a child from school (Alvarez and Schieflbein, 2007). High rates of malnutrition are associated with a 50% larger probability of dropping out of school, and about 100% higher grade repetition rates (Martínez and Fernández, 2007).

Pre-primary education covers children aged 4 to 6, while the first grade of primary school typically started at age 7 (though these thresholds are not strict). While primary education has been compulsory since 1985, parents have to pay a small fee (between \$0.60 and \$5 for tuition) to cover operational costs related to running the school (e.g., electricity bills). Pre-primary education is compulsory as well, though not enforced due to low coverage.

The signing of the Peace Accords in 1996 put an end to 36 years of harsh civil conflict. As part of the Accords, the national government agreed to expand basic education and health services in rural and indigenous areas with inadequate coverage. The government aggressively expanded access to public pre-primary education, increasing the number of public pre-primaries from approximately 5,300 in 1998 to 11,500 in 2005 (Figure 1). Pre-primaries were usually constructed as annexes to primary schools.

According to government officials we interviewed, the selection of beneficiary communities can be described as a two-step process. In the first step, agents at the departmental offices of the Ministry of Education identified eligible communities as those that had enough children without access to pre-primary education. As a result, lists of eligible communities were constructed at the regional level. In the second step, final decisions were made at the central level. The procedures were *ad hoc* and no strict guidelines were applied to make decisions on beneficiary schools.

## 3 Data

We use school-level administrative data obtained from the Ministry of Education for all primary schools from 1995 until 2008. At the beginning of the school year, each operating educational establishment (for example, pre-primary, primary or high school) must send information on initial enrollment by grade, age and gender. Non-response is very low (less than 1%). Using this information it is possible to construct a panel at the community-cohort level. Unfortunately, we do not have access to data on measures of school readiness and learning performance among students in the sample. Counting with data on these dimensions would have allowed us to provide a more comprehensive analysis of the short- and medium-run effects of attending a pre-primary. Additionally, we have data on private pre-primaries and their exact location for 2005 but not for prior years.

We have also gained access to comprehensive data from the 1994 and 2002 Population Census. These data contain basic socioeconomic characteristics at the household and individual level. Importantly for our purposes, the geographic location of the household is identified at the community level, and it is possible to match communities from the 1994 to the 2002 Population Census. Additionally, the National Statistics Institute provides geo-coordinates for communities in the 2002 Population Census.

# 4 Research design

## 4.1 Sample construction

We focus on a sample of rural communities that had a primary school during the period of analysis. We exploit variation in access to pre-primaries across cohorts within communities to identify the impacts on enrollment and progression rates.

To be selected into the estimation sample, a rural community has to meet 4 conditions. First, it must have a population of between 300 and 2,500 in 2002. This condition ensures that rural communities under analysis are large enough to have a pre-primary school. Second, it needs to have one primary school operating during the 1995 to 2008 period. This makes it possible to measure impacts on enrollment and progression in primary school. Third, the community should not have other communities located within a 0.5 kilometers radius from its center This restriction reduces the likelihood that the opening of a pre-primary in one community might induce children from neighboring communities to start attending pre-primary education there.[5] Finally, the community must not have a pre-primary by 1997. This condition ensures that selected communities are homogeneous in terms of no baseline access to pre-primary education and allows us to inspect for the

[5]This type of spill-over effects would introduce a downward bias in the estimates. Their presence is minimized in this context because of the rugged geography and poor infrastructure that make it difficult for children in non-beneficiary communities to attend pre-primaries in treatment communities. Anecdotal evidence also suggests that cultural factors related to a strong sense of belonging to the community reduce the possibility of seeking services elsewhere. Finally, the fact that young children would have to travel by themselves to attend pre-primary school outside their community makes this possibility less likely. Still, in section 6.1 we explore whether the main findings could be affected by spill-over effects.

existence of pre-trends.

We define treatment status at the community-cohort level. Children from a given cohort are considered treated if a pre-primary existed in their community when they were aged 6. We measure impacts of access to pre-primary education on two outcomes: enrollment and adequate progression rates. We restrict our sample to include cohorts born between 1988 and 1996 to observe outcomes during the whole relevant age range (7-12) for each included cohort. This restriction ensures that differences in estimated effects by age are not driven by heterogeneity in cohort composition, but rather by average dynamic effects of pre-primary access.

For both measures, the denominator corresponds to population counts by community-cohort-age. To construct these counts, we use data on population by community and age from the 1994 and 2002 Population Census and assume that cohort sizes remain constant over time. For example, the number of students aged 6 in 2000 in a given community is assumed to be equal to the number of students aged 8 in 2002 in that community. We compute counts for cohorts 1988-1989 using the 1994 Population Census. For the 1990-1996 cohorts we use the 2002 Census. We favor the use of the 2002 Census because it is closer to the period of analysis (1995-2008). However, the assumption of constant cohort sizes by community over time becomes problematic when focusing on children aged 13 and older in the 2002 Population Census: in these rural communities students might migrate to attend secondary education. This is why we compute cohort sizes for the 1988 and 1989 cohorts from the 1994 Population Census.[6]

The numerator of the enrollment rate is the count of primary school students by community-cohort-age. The numerator for the progression rate corresponds to the number of students progressing adequately given their age. For example, for those aged 8, this includes all students in second or higher grades. The unit of observation of the resulting data set is at the community-cohort-age level. This structure of the data makes it possible to estimate impacts at ages 7 to 12, though we mainly concentrate on impacts at age 12 – the medium-run effects.

Our empirical analysis is based on comparing changes in outcomes between later (exposed) and earlier cohorts in treatment communities versus changes in comparison communities. Note that our analysis is performed at the community level and that the treatment indicator equals one if a cohort-community had access to a pre-primary and zero if not. Moreover, in our baseline analysis we restrict the sample to communities that did not have access to a pre-primary prior to the pre-primary construction program. Consequently, pre-primary enrollment rates for all treatment and comparison communities are expected to

[6]The results presented below remain robust when using the 2002 Population Census for these cohorts instead. Dropping them from the sample also leaves our findings unaltered.

be close to 0 prior to the construction program.

For both sets of communities, we have information for pre-treatment (earlier) cohorts; that is, children born between 1988 and 1991. Table 1 reports summary statistics. The upper panel of this table documents that pre-treatment enrollment and progression rates are significantly higher for schools in the treatment as compared to those in comparison communities. It also shows that enrollment rates are high and close to 1, whereas adequate progression rates are much lower and decreasing with age. The lower panel shows that while there are some differences in observed socio-demographic variables between both groups they tend to be small. These differences in pre-treatment outcomes and baseline characteristics can be accommodated in the difference-in-differences framework we adopt.

## 4.2 Empirical strategy

In order to estimate the effects of expanding access to pre-primary services in rural communities on the corresponding primary enrollment and progression rates we employ a difference-in-differences model with trimming and propensity score re-weighting. We define the unit of observation at the community-cohort-age level. Outcome variables are defined at this level and the treatment dummy is defined at the community-cohort level. The baseline specification is a parsimonious difference-in-differences model that exploits the variation in access to pre-primary services within communities across cohorts to identify impacts. Specifically, we estimate the following baseline equation:

$$Y_{cba} = \beta Access_{cb} + \lambda_c + \delta_b + \varepsilon_{cba} \tag{1}$$

where $Y_{cba}$ is the outcome of interest for the community $c$, birth-cohort $b$ at age $a$. The coefficient of interest ($\beta$) estimates the intention-to-treat (ITT) parameter measuring the average effect of having access to a pre-primary school on the outcomes of interest. We include fixed-effects by community ($\lambda_c$) and birth cohort ($\delta_b$) to account for time-invariant heterogeneity in the treatment and comparison communities. Since errors can be expected to be correlated within community over different cohorts, we cluster standard errors at the community level.

The key identifying assumption of the baseline specification is that, in the absence of the treatment, outcomes in communities where pre-primaries were opened would have evolved similarly as in those communities where that did not happen. If departments targeted in the pre-primary expansion program benefited from other public programs that could potentially affect the outcomes of interest, the estimated effects yielded by (1) would be biased. Moreover, if shocks in outcomes at the department (or municipality) level were correlated with pre-primary openings (for example, if departments with worse trends in outcomes were targeted for compensatory programs such as pre-primaries expansions), our estimates would also be biased.[7] We can tackle these potential threats to identification by extending the baseline specification to include department dummies interacted with a linear time trend term. These interactions account for changes in outcomes in departments over time that may bias the estimates obtained from the original specification. In this case, we estimate the following equation:

$$Y_{cba} = \beta Access_{cb} + \lambda_c + \delta_b + t\gamma_d + \varepsilon_{cba} \tag{2}$$

where $t$ corresponds to years elapsed after 1995, $\gamma_d$ represents department (or municipality) indicator variables and, thus, the interaction would capture department-specific trends over time.

We additionally exploit data from the 1994 and 2002 Population Census on the prevalence of households with running water, concrete floor and whose heads have no education and are indigenous. Using these data we generate community-year covariates by linear extrapolation. We add these covariates to the baseline specification to control for observable changes at the community-level over time. The estimated model is:

$$Y_{cba} = \beta Access_{cb} + \lambda_c + \delta_b + t\gamma_d + \varphi X_{ct} + \varepsilon_{cba} \tag{3}$$

where $X_{ct}$ is a vector of covariates defined at the community-year level.

In all specifications we apply trimming and propensity score re-weighting techniques in order to increase the similarity in observable dimensions of communities in the treatment and comparison groups. We implement this method by first predicting the probability, using a logit regression, that a pre-primary is opened in a community as a function of community characteristics. Then, we drop communities with probability of treatment above the 85th percentile or below the 15th percentile. Finally, we re-weight observations in the comparison group applying a factor of $\frac{PScore}{1-PScore}$ where $PScore$ refers to the estimated probability of treatment, and estimate the basic specification in the trimmed and re-weighted sample.[8]

The specifications described above make it possible to estimate ITT parameters of the impact of opening pre-primaries in rural communities on primary school progression. To estimate the effect of attending a pre-primary – the treatment-on-the-treated (ToT) parameter – we compute the pre-primary take-up rate for each cohort at age six and use an instrumental variable approach to tackle the endogenous nature of this attendance rate.

[7] There are 22 departments and 332 municipalities in Guatemala.

[8] See online appendix for more details about the sample distribution when weighted and unweighted as well as robustness of results to different trimming and weighting specifications.

In the first stage, we regress the take-up rate on the pre-primary access indicator defined at the community-cohort level. In the second stage, we estimate the model described above but specifying the fraction of children attending the pre-primary at the cohort-community level as the main independent regressor and instrumenting this variable with the pre-primary access dummy.[9]

# 5 Results

## 5.1 Main impacts

We first examine if and how the opening of pre-primaries in rural communities affects primary school progression by age 12. Table 2 presents the corresponding ITT estimates of the average effect of interest. The baseline specification described above is presented in column (1). We find no statistically significant effects associated with having access to a pre-primary on enrollment rates (Panel A). However, we do find statistically significant positive effects on adequate school progression (Panel B). The results indicate that opening a pre-primary in these rural communities increased by approximately 2.4 percentage points the proportion of children in grade 6 by age 12.

In columns (2) to (4), we test the robustness of these results using alternative specifications. Column (2) presents results when controlling for department-linear time trends and column (3) when controlling for differential municipality-linear time trends. We exploit the matched Census 1994 and 2002 data and control for time-varying controls in column (4). The baseline estimates are robust to these alternative specifications.

We proceed by examining the ToT average effects of attending a pre-primary on primary school progression in Table 3. Because the average take-up rate is about 60 percent for a given cohort at age 6, the estimated effects are roughly 1.67 times the ITT estimates of Table 2. The results indicate no significant effects on enrollment rates, but show a positive statistically significant effect on adequate progression of about 4.2 percentage points. As before, the results are robust to changes in the empirical specification used.

Expanding access to pre-primary education seems to improve adequate progression, while not changing enrollment rates in primary school. Taken together this evidence suggests that repetition rates are reduced by this intervention. This is important in Guatemala, where repetition rates in early grades are high (about 30 percent in first grade and 10-15 percent in later grades). However, the magnitude of the effect on adequate progression (about 2.4 percentage points) seems fairly modest when considering that only

[9] To estimate treatment-on-the treated estimates we assume that pre-primary attendance in non-beneficiary communities is zero. In the case this is not true, our estimates of the effect of attending a pre-primary would be downward biased and, hence, the estimates can be interpreted as a lower bound of the true effect.

12.3 percent of children are in the expected grade by age 12. On the other hand, the lack of effects on primary enrollment could have been expected given that the overall baseline coverage is high. Still, if improvements in adequate progression are signaling that children are better prepared to the challenges faced in primary education, the intervention may produce changes in drop-out in lower grades.

The positive effects on adequate progression are in line with Berlinski et al. (2009) who document sizable effects of pre-primary attendance on test scores in primary school in Argentina. In Uruguay, attending preschool was associated with a large increase in the probability of attending school at age 15 of 27 percentage points (Berlinski et al., 2008). However, as mentioned above, extrapolating these positive findings to rural contexts in poorer countries such as Guatemala is not straightforward. First, stimulation of skills conducive to high academic achievement is likely to be significantly worse in Guatemala's rural households than in those of Argentina and Uruguay. For example, average mothers' education in the Uruguay study amounted to 9.8 years, whereas in the communities included in our study average education for women aged 20 to 40 is only 3.2 years. Second, reports from several sources suggest that the quality of pre-primary education in Guatemala is low, and this could attenuate any positive benefits related to attending a pre-primary (UNICEF, 1996; Rubio et al., 2001). Finally, Guatemala has one of the highest levels of child chronic malnutrition in Latin America (Pan American Health Organization, 2008), which could potentially offset any positive effects associated to pre-primary attendance.

### 5.2 Further results

Figure 2 presents point estimates and 95% confidence intervals for the effects of pre-primary access at different ages. Results suggest negative effects of pre-primary access on primary enrollment at age 7 (p-value 0.082). This is consistent with evidence from Berlinski et al. (2008) for Uruguay. It can be the case that children aged 7 who are not adequately prepared for first grade are kept longer in pre-primary whereas their counterparts in communities without pre-primaries are enrolled in primary school. These negative effects on enrollment vanish at later ages suggesting that the mechanism at work here is delayed entry to primary school induced by pre-primary access. Regarding adequate progression, statistically significant effects only arise by age 12 suggesting accumulated effects over time.

Do effects differ across subgroups of the population? Identifying heterogeneous impacts can provide useful guidance on targeting pre-primary expansion programs to maximize expected effects. For clarity, Table 4 presents only estimated effects for the baseline specification and that controlling for department-linear time trends (columns 2 and 4). Panel A shows results from a regression where the relevant explanatory variables are a treatment indicator and the interaction of this variable with a dummy for girls. Similarly, in Panel B the treatment indicator is interacted with a dummy that signals that the school has a higher enrollment than the median school. Panels C and D report results from analogous regressions when splitting the sample by prevalence of indigenous population and by the share of adults with no education. Finally, in Panel E we explore the extent to which malnutrition could be attenuating the positive effects of pre-primary attendance. We matched data at the municipality level from the Second National Census of Height of 2001 (*Segundo Censo Nacional de Talla*). They measured the height for all first grade students aged 6 through 9, constructing an indicator of prevalence (%) of stunted children at the municipality level.[10]

There are several noteworthy findings. First, we find that the effect of pre-primary attendence does not appear to differ significantly by gender (Panel A). Documenting similar effects between genders is in line with the aforementioned studies for Argentina and Uruguay (Berlinski et al., 2008; Berlinski et al., 2009). However, evidence from the United States document larger effects of preschool programs for girls (Oden et al., 2000; Anderson, 2008; Cascio, 2009). Second, we do not find consistent differences in effects on enrollment when classifying communities by percentage of indigenous population, share of adults with no education or prevalence of chronic malnutrition (Panels C to E); .

Similar to the effects on enrollment, when considering progression we find no statistically significant differences for populations with higher proportions of indigenous population or with a high prevalence of chronic malnutrition. However we do find that the positive effects of pre-primary access are attenuated in communities where a high share of adults has no education. This may be due to several factors associated with higher educational attainment in the community. For example, pre-primary take up rates may be higher in communities with higher education since it may be valued more by parents. Nevertheless, this does not seem to be the case since we do not find relevant differences in take-up rates between communities with high and low shares of adults with no education.[11] Alternatively, these estimates may reflect positive interactions between pre-primary attendance and parental education, or the fact that teacher quality is likely to be higher in communities with higher average education.

# 6 Robustness

For robustness, we address several potential challenges to our difference-in-differences estimates. To assess the validity of the basic identification assumption, we then test for the

[10]In our sample, the mean prevalence is around 50%.

[11]The average take-up is 60.1% and 57.4% for low and high share of adults with no education (p-value: 0.237).

existence of differential pre-trends between treatment and comparison communities.

## 6.1 Challenges to the identification strategy

We verify the sensitivity of the main results with respect to five potential concerns about the identification strategy. First, the effect of pre-primary access on enrollment and progression my be spurious in the sense that communities that had a pre-primary built may have also received additional benefits at the same time which are unknown to the researcher. We tackle this concern by examining whether pre-primary access had heterogeneous effects by the average pre-primary take-up rate in the community.

Table 5 presents results for the interaction between the pre-primary access dummy and dummies for low, medium and high pre-primary take-up.[12] If results were spurious (i.e. driven by other factors than pre-primary attendance), we would expect to find similar effects across groups. The results strongly reject this possibility: the effect of pre-primary access is strongly increasing in average pre-primary take-up for the community.

A second concern is whether the results are sensitive to the selection of the comparison group. Even though we are applying trimming and propensity-score re-weighting, there could be unobservable factors driving the selection of beneficiary communities that would not be taken into account and could be driving our results. We address this issue by restricting the sample to beneficiary communities and exploiting the variation of timing to identify the effect or pre-primary access. The results are presented in columns (3) and (4) of Table 6, where both the OLS and IV estimates are presented. Noticeably, these estimates are not statistically different to those from our baseline estimates (in columns (1) and (2)).

A third challenge is whether the results are sensitive to the exclusion of communities that already had a pre-primary by 1997. As discussed above, the focus on communities without a pre-primary school by 1997 makes it possible to use a more homogeneous sample and inspect differential pre-treatment trends for beneficiary communities. We now verify if the estimates remain robust when relaxing this restriction. The results, presented in columns (5) and (6), do not show significant effects on enrollment rates, but point to positive and significant impacts on adequate progression rates.[13]

A fourth challenge relates to the evidence suggesting that measurement error is present

[12] Beneficiary communities are classified by average pre-primary take-up rate by terciles. The average pre-primary take-up rates for the low, medium and high take-up communities are 31.2%, 54.3% and 90.3%.

[13] Additionally, Table A6 in the online appendix shows that results are qualitatively similar when using alternative measures of pre-primary availability instead of using an indicator for access to a pre-primary in the community. The first measure equals the number of pre-primaries in the community divided by the population of pre-primary aged children in 2002. The second measure equals the number of pre-primaries in the community divided by the population of pre-primary aged children in each year. Additional details are presented in the online appendix.

in the outcome variables, notably in enrollment rates. About half of the communities in the sample present average enrollment rates for the pre-treatment period above 1. A substantial number of communities with average enrollment rate above one could arise if mean coverage is close to 1 and there is some noise in this measure because the ratio is constructed from two different data sets. However, the fact that the average rate across communities is above 1 at certain ages does suggest that there is some over-reporting of students enrolled or under-reporting of the number of children in the communities in the sample. If measurement error in these variables is classical and uncorrelated with the introduction of pre-primaries, it will only reduce precision. However, this assumption cannot be tested. Hence, we verify if we can replicate the main results by focusing on those communities that in the pre-treatment period had average enrollment below 1. Results, presented in columns (7) and (8), once again indicate no effects on enrollment rates, but point to positive and significant impacts on adequate progression rates.[14]

Finally, in constructing the main sample for the analysis, we selected communities that did not have a neighboring community within a 0.5 kilometers radius from its center. The motivation for this restriction was to reduce the possibility of spill-over effects that may introduce biases in the estimation. For example, if children in comparison communities attended the newly constructed pre-primary in a beneficiary community this would induce a downward bias in our estimates. Similarly, if the opening of a pre-primary in a community induces changes in the composition of children attending the primary school in that community (because children in neighboring communities start attending there), this will also bias our estimates.

We consider that setting the described minimum-distance restriction could be sufficient to reduce the possibility of spill-over effects in this context based on the following reasons. To start with, it has been recognized that parents usually desire to have their young children attending educational centers located close from home (e.g. see Paes de Barros et al., 2011 for evidence on Brazil). Consistent with this view, 97% of children ages 4 to 6 years old attending a pre-primary in rural Guatemala walk 30 minutes or less to attend an educational center (Living Standards Measurement Survey, 2006). Moreover, the reluctance of parents to have their children attending pre-primaries outside their communities may be heightened in this context due to cultural factors related to a strong

[14]One additional concern is the possibility that a community that does not have a public pre-primary might have a private one (this could attenuate our estimates). As mentioned above, we have data on private pre-primaries and their exact location for 2005 but not for prior years. Consequently, we cannot assess whether the expansion in public pre-primaries affected the supply of private pre-primaries in our sample. Still, we believe that it is unlikely that private supply will play a relevant role in this context because of the overall little demand for private pre-primaries in rural Guatemala. This is well exemplified in the fact that only 1.2% of communities in the comparison group and 0.8% of treatment communities had a private pre-primary in 2005.

sense of belonging to the community (Rao and Walton, 2004). Additionally, distance *per se* may not be the main barrier for educational enrollment decisions but rather related safety considerations (Lloyd et al., 2005). This issue can be particularly salient for young children in this context given that the civil strife that took place in Guatemala between 1960 and 1996 affected predominantly rural areas (Chamarbagwala and Morn, 2011).

On the other hand, the mentioned 0.5 kilometers minimum-distance restriction applies to the linear distance to *any* community. In rural Guatemala the population typically lives in mountainous, rugged areas with poor access to infrastructure or public transportation. Therefore, the linear distance would dramatically underestimate a simple calculation regarding how long it would take to a child to walk to neighboring community to attend a pre-primary. Also, many communities in rural Guatemala are very small and they can be better interpreted as groups of households. For example, the community in the 25th percentile in the population distribution of the 2002 Census contained only 40 individuals. Finally, we note that the possibility of spill-over effects may be less of an issue for girls compared to boys, as existing empirical evidence suggests that the distance to an educational establishment may be a stronger barrier for girls (Burde and Linden, 2012). Hence, if spill-over effects are at play they will operate predominantly for boys, and, consequently, we should observe differences in effects across genders. However, Table 4 shows that effects are similar across genders suggesting a limited role for spill-over effects.

Notwithstanding these considerations, we empirically explored the robustness of the main findings to the minimum-distance restriction. In particular, we estimated heterogeneous treatment effects across communities that vary in their distance to its closest neighboring community. If spill-over effects play a role in the estimated effects, we should expect differential treatment effects across distance to neighboring communities. In practice, we divided the sample of rural communities under analysis based on whether the distance to the neighboring community was above or below the median distance (0.75 kilometers). We found that effects are qualitatively similar across the two sub-samples suggesting that spill-over effects are not affecting the main findings in the study (results are presented in Table 7).

### 6.2 Testing for differential pre-intervention trends

Identification of the parameter of interest relies on the assumption that, in absence of the treatment, outcomes in the intervened primary schools would have evolved similarly to those from the comparison group. Although this assumption cannot be tested directly, we can provide some evidence on its plausibility by studying pre-intervention trends (Heckman and Hotz, 1989). To this end, we perform the following falsification tests. First, we restrict the sample to only the pre-construction cohorts (1988 to 1991 cohorts) - that is, those that could not have been exposed to pre-primary education. We create a placebo dummy that equals one for the 1990 and 1991 cohorts in the beneficiary communities, and zero otherwise. The placebo indicator would measure any differences in trajectories for the variable of interest prior to the construction of pre-primaries. The results for this specification are presented in columns (1) and (2) of Table 8. The point estimates for the placebo indicator are close to zero and not statistically different to zero at conventional levels.

The second test takes advantage of the entire sample and, in addition to our pre-primary access dummy, we add a placebo indicator that takes the value of 1 for cohorts born one or two years prior to the first exposed cohort in beneficiary communities. For example, if the 1995 cohort was the first to be exposed to a pre-primary in a certain community, then the 2-year prior placebo indicator would take the value of 0 for cohorts 1990 to 1992, 1 for the 1993 and 1994 cohorts and 0 for cohorts 1995 and 1996. Columns (3) through (6) present the results for this test. The point estimates on the placebo indicators are similar to those obtained in the first two columns, albeit more precisely estimated given that we are including the entire sample. Note that the estimates for pre-primary access remain unchanged.

Finally, we use the 1994 and 2002 Population censuses to test whether the change in community characteristics between these two points in time is different between beneficiary and comparison communities, which could potentially bias our results. We present the estimates for this test in Table 9. We find that the change in several socio-economic characteristics is, in almost all cases, not significantly different between beneficiary and comparison communities, again suggesting that our main identification assumption is valid.

# 7 Concluding remarks

Many developing countries have recently expanded (or are considering to expand) pre-primary coverage as an instrument to stimulate human capital accumulation. Evidence from the United States, Uruguay and Argentina suggests that these investments yield large returns (Cascio, 2009; Berlinski et al., 2008; Berlinski et al., 2009). But this evidence cannot be easily extrapolated to poorer countries, where the potential for policy interventions in this domain is much greater.

We have contributed to filling this gap in the literature by exploiting a large-scale expansion in pre-primary coverage in rural Guatemala from 1998 to 2005. Combining administrative and population census data in a difference-in-differences framework, we have estimated the medium-run impacts of opening a pre-primary on enrollment and adequate progression in primary school by age 12. We find positive effects on the proportion of students progressing adequately by this age. These effects are heavily concentrated among more educated communities, and among those with higher pre-primary take-up rates.

The average positive effects on adequate progression amount to about 2.4 percentage points. These effects appear modest when compared to the average adequate progression rate of 12.3%. Hence, they suggest that tackling high rates of repetition will require significant complementary actions including improvements in the quality of primary education. The take-up rates of about 60% that we document suggest that programs seeking to increase attendance may be a valuable policy option. For example, adding to existing cash transfer programs a condition on pre-primary attendance might be an inexpensive way to improve the impacts generated by both social programs. Additionally, our findings suggest that the effectiveness of a large scale expansion in pre-primary may be small if the quality of instruction is low. This is not only relevant for low- and medium-income countries, but also in the United States, where universal early childhood education is a policy goal.

One limitation of our paper is that we are unable to examine impacts on test scores, because of data limitations. Nonetheless, estimating impacts on enrollment and adequate progression is interesting in its own right for several reasons. First, under the assumption that students below a specific threshold are assigned to repeat the grade, increases in adequate progression signal improvements in learning, at least for students at the margin of failing. Second, evidence from the United States suggests that the impacts of preschool attendance on test scores tend to fade out rapidly, whereas positive effects on school progress and other non-cognitive outcomes typically remain (or even arise only) in adolescence or later in life (Almond and Currie, 2011). Impacts on adequate promotion can plausibly increase years of schooling. Third, in many developing countries high rates of repetition, school dropout and re-entry increase public spending and generate problems associated with heterogeneity in the classroom. Hence, evaluating programs that can alleviate these issues is also desirable.

Future research using individual-level data and experimental designs may provide more definitive answers about the impacts of expanding pre-primary coverage. Yet evidence on the impacts of large-scale expansions in access to preschool education will most likely come from non-experimental designs exploiting significant policy shifts like the one examined here. Together, both approaches will be informative of effective ways to promote human capital accumulation in less developed countries.

## References


Almond, D., and J. Currie (2011). "Human Capital Development before Age Five." in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 4B, Chapter 15: 1315-1486. North Holland: Elsevier.

Alvarez, H., and E. Schieflbein (2007). "Informe Integrado del Sector Educación." Washington, DC: Inter-American Development Bank. Mimeographed document.

Anderson, M. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103(484): 1481-1495.

Baker, M., J. Gruber, and K. Milligan (2008). "Universal Childcare, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy* 116(4): 709-745.

Barros, R., M. de Carvalho, S. Franco, R. Mendonça and A. Rosalém (2011). "A Short-Term Cost-Effectiveness Evaluation of Better-Quality Daycare Centers." Working Paper IDB-WP-239. Washington, DC: Inter-American Development Bank.

Berlinski, S., S. Galiani, and P. Gertler (2009). "The Effect of Preprimary Education on Primary School Performance." *Journal of Public Economics* 93(1-2): 219-234.

Berlinski, S., S. Galiani, and M. Manacorda (2008). "Giving Children a Better Start: Preschool Attendance and School-Age Profiles." *Journal of Public Economics* 92(5-6): 1416-1440.

Black, S., P. Devereux, and K. Salvanes (2005). "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital." *American Economic Review* 95(1): 437-449.

Calderón, V., and M. Urquiola (2006). "Apples and Oranges: Educational Enrollment and Attainment across Countries in Latin America and the Caribbean." *International Journal of Educational Development* 26(6): 572-590.

Cascio, E. (2009). "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools." NBER Working Paper 14951. Cambridge, United States: National Bureau of Economic Research.

Chamarbagwala, R., and H. Moran (2011). "The Human Capital Consequences of Civil War: Evidence from Guatemala." *Journal of Development Economics* 94(1): 41-61.

Duflo, E. (2001). “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment.” *American Economic Review* 91(4):795-813.

Engle, P., L. Fernald, H. Alderman, J. Behrman, C. O’Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, S. Iltus, and the Global Child Development Steering Group (2011). “Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries.” *The Lancet* 378 (9799): 1339-1353.

Glewwe, P. (1999). “Why Does Mother’s Schooling Raise Child Health in Developing Countries? Evidence from Morocco.” *Journal of Human Resources* 31(1): 134-159.

Heckman, J., and V. Hotz (1989). “Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training.” *Journal of the American Statistical Association* 84(408): 862-874.

Lloyd, C., C. Mete, and Z. Sathar (2005). “The Effect of Gender Differences in Primary School Access, Type, and Quality on the Decision to Enroll in Rural Pakistan.” *Economic Development and Cultural Change* 53(3): 685-710.

Luo, R., L. Zhang, C. Liu, Zhao, Y. Shi, S. Rozelle, and B. Sharbono (2012). “Behind Before They Begin: The Challenge of Early Childhood Education in Rural China.” *Australasian Journal of Early Childhood* 37(1): 55-64.

Martín, T., and F. Juárez (1995). “The Impact of Women’s Education on Fertility in Latin America: Searching for Explanations.” *International Family Planning Perspectives* 21(2): 52-80.

Martínez, R., and A. Fernández (2007). “Análisis del Impacto Social y Económico de la Desnutrición Infantil en América Latina. Resultados del Estudio en Centroamérica y República Dominicana.” Santiago, Chile: CEPAL, División de Desarrollo Social. Mimeographed document.

Oden, S., L. Schweinhart, and D. Weikart (2000). “Into Adulthood: A Study of the Effects of Head Start.” Ypsilanti, United States: High/Scope Press.

Pan American Health Organization (2008). “Malnutrition in Infants and Young Children in Latin America and the Caribbean: Achieving the Millennium Development Goals.”Washington, DC: Pan American Health Organization.

Rao, V., and M. Walton (2004). “Culture and Public Action.” California: Stanford Social Sciences.

Rodríguez, M. (2001). “Percepciones sobre la Educación: Un Estudio Cualitativo y Multiétnico en Guatemala.” Guatemala Poverty Assessment Program, Technical Paper 4, Part A. Washington, DC: World Bank.

Rubio, F., E. Fernando and R. Chávez (2001). “An Evaluation of the Early Childhood Education and Preschool Program Implemented by Niños Refugiados del Mundo: Classroom Implementation and Community Participation.” Improving Educational Quality (IEQ) Project. Washington, DC, United Sates: American Institute for Research. Mimeographed document.

Todd, P. (2007). “Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated.” in T. Schultz and J. Strauss (eds.) *Handbook of Development Economics*, Volume 4, Chapter 60: 3847-3894.

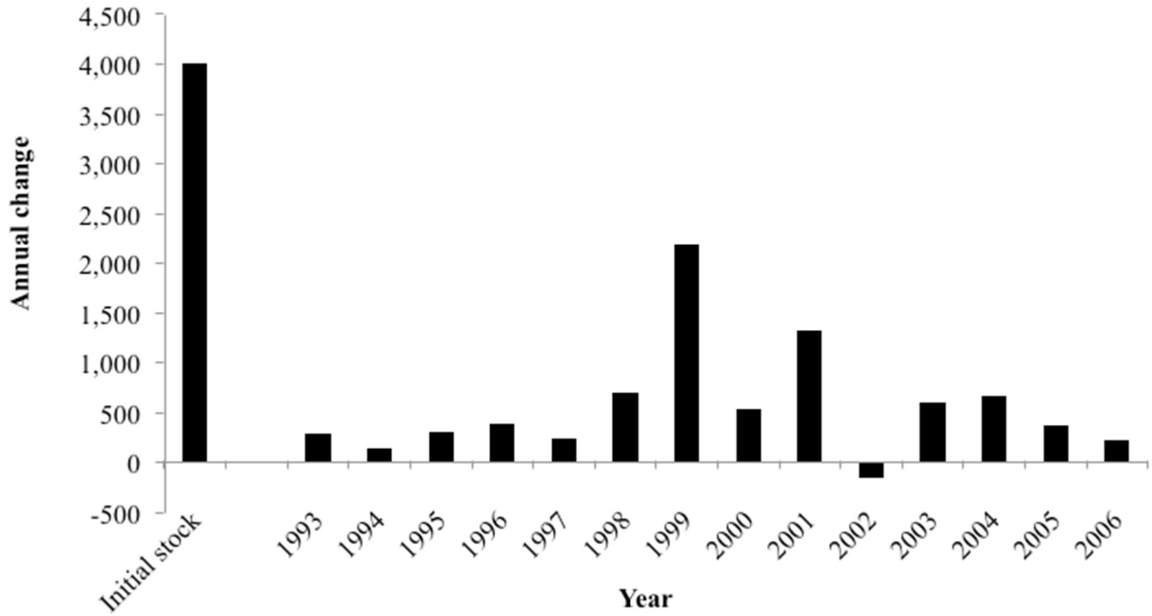
UNESCO (2006). “Preprimary Education: A Valid Investment Option for EFA.” UNESCO Policy Brief on Early Childhood, Number 31. Paris, France: UNESCO.

UNICEF (1996). “Proyecto Centro de Aprestamiento Comunitarios en Educación Preescolar CENACEP.” New York: UNICEF. Mimeographed document.

Wong, A., R. Luo, L. Zhang, and S. Rozelle (2013). “The impact of vouchers on preschool attendance and elementary school readiness: a randomized controlled trial in rural China.” *Economics of Education Review* 35:53-65.

World Bank (2012). World Development Indicators. Washington, DC: World Bank.

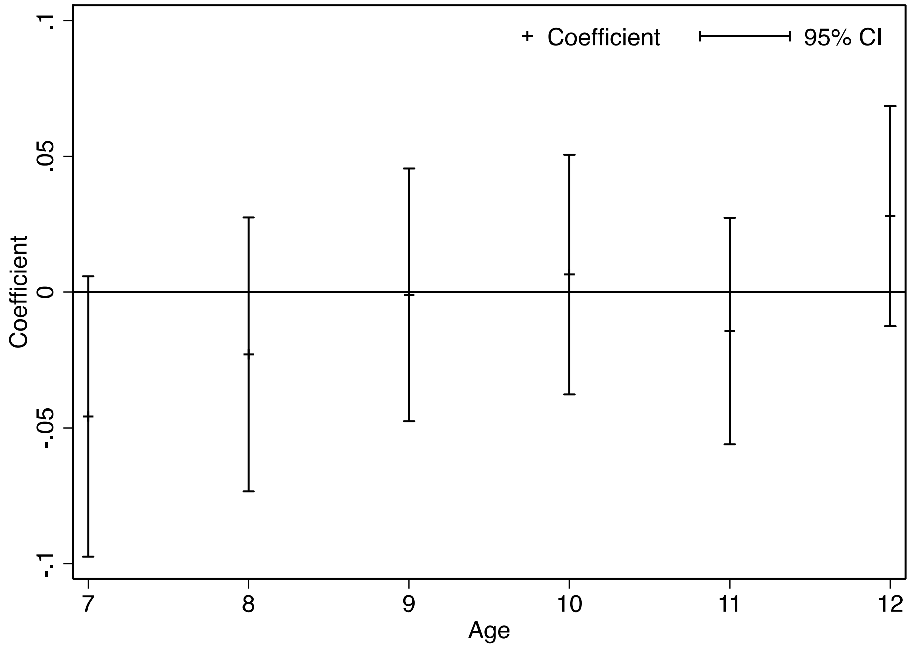
**Figure 1: Annual change in stock of pre-primaries**



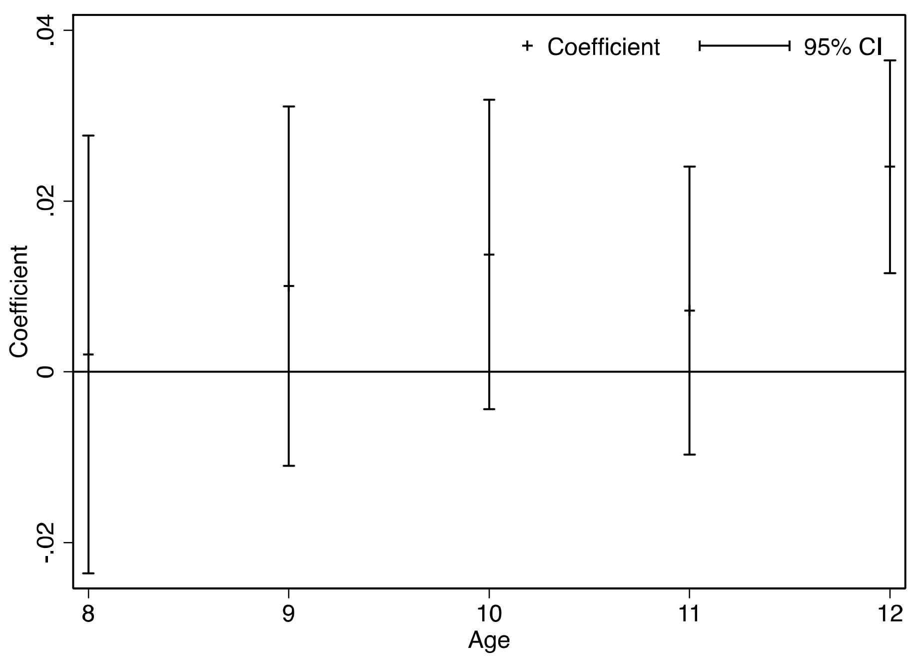

*Notes*: This figure presents the annual change in the stock of pre-primaries in Guatemala. For comparison purposes, the leftmost bar plots the stock of pre-primaries in 1992.

**Figure 2: Impact of expansion of pre-primary by age**

a) % of students enrolled in primary school




b) % of students progressing adequately




*Notes*: The graphs present coefficients and 95% confidence intervals of the effects of pre-primary access on enrollment and adequate progression. The unit of observation is community-cohort. Separate regressions are estimated by age. All regressions are estimated applying trimming and propensity score re-weighting with cohort and community fixed effects. Standard errors are clustered at the community level.

## Table 1: Sample descriptive statistics, pre-construction cohorts

| | Treatment | Comparison | Raw Difference | Adjusted Difference |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A - Primary school outcomes** | | | | |
| % Enrolled at ages 7-12 | 1.021 | 0.969 | 0.052 | 0.056 |
| | [0.005] | [0.004] | [0.007]*** | [0.009]*** |
| % Progressing adequately at ages 8-12 | 0.227 | 0.203 | 0.023 | 0.010 |
| | [0.003] | [0.002] | [0.003]*** | [0.004]** |
| % Enrolled at age 8 | 1.118 | 1.039 | 0.078 | 0.090 |
| | [0.014] | [0.011] | [0.018]*** | [0.023]*** |
| % Enrolled at age 12 | 0.892 | 0.854 | 0.038 | 0.029 |
| | [0.011] | [0.009] | [0.014]*** | [0.018] |
| % Progressing adequately at age 8 | 0.354 | 0.324 | 0.030 | 0.014 |
| | [0.008] | [0.006] | [0.010]*** | [0.012] |
| % Progressing adequately at age 12 | 0.123 | 0.109 | 0.013 | 0.005 |
| | [0.003] | [0.003] | [0.004]*** | [0.005] |
| **Panel B - Community characteristics** | | | | |
| % Indigenous | 0.348 | 0.344 | 0.004 | -0.013 |
| | [0.018] | [0.015] | [0.023] | [0.030] |
| % No education | 0.270 | 0.309 | -0.038 | -0.004 |
| | [0.004] | [0.004] | [0.006]*** | [0.006] |
| Average age | 21.806 | 21.749 | 0.057 | 0.205 |
| | [0.104] | [0.084] | [0.134] | [0.171] |
| % Household has kitchen | 0.848 | 0.832 | 0.015 | -0.007 |
| | [0.007] | [0.006] | [0.009]* | [0.011] |
| % Household has running water | 0.597 | 0.575 | 0.022 | 0.026 |
| | [0.016] | [0.012] | [0.020] | [0.025] |
| % Household has concrete floor | 0.350 | 0.274 | 0.076 | 0.015 |
| | [0.011] | [0.008] | [0.014]*** | [0.017] |
| Community size | 767.609 | 688.364 | 79.244 | 14.431 |
| | [17.299] | [12.464] | [21.319]*** | [25.430] |
| Number of communities | 516 | 857 | 1,373 | 960 |

*Notes*: This table presents statistics and estimated differences between communities that eventually had access to a pre-primary (Treatment) and those that did not (Comparison). Panel A reports statistics generated from school administrative records for pre-construction cohorts (those born between 1988 and 1991). Panel B presents statistics obtained from the 2002 Population Census. Columns 1 and 2 present means, column 3 presents the raw difference across groups. Column 4 presents differences after applying trimming and propensity-score re-weighting (described in section 4.2). Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 2: Impacts of pre-primary access, ITT estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Panel A - Dependent variable: % Enrolled in primary at age 12** | | | | |
| Pre-primary access (=1) | 0.030 | 0.028 | 0.027 | 0.027 |
| | [0.021] | [0.021] | [0.022] | [0.021] |
| R-Squared | 0.022 | 0.025 | 0.061 | 0.027 |
| Number of observations | 8,543 | 8,543 | 8,543 | 8,543 |
| **Panel B - Dependent variable: % Progressing adequately at age 12** | | | | |
| Pre-primary access (=1) | 0.024 | 0.024 | 0.020 | 0.024 |
| | [0.006]*** | [0.006]*** | [0.007]*** | [0.006]*** |
| R-Squared | 0.061 | 0.068 | 0.102 | 0.068 |
| Number of observations | 8,543 | 8,543 | 8,543 | 8,543 |
| Number of communities | 960 | 960 | 960 | 960 |
| Department-linear time trends | | Yes | | Yes |
| Municipality-linear time trends | | | Yes | |
| Time-varying controls | | | | Yes |

*Notes:* Pre-primary access is a dummy variable that equals 1 for cohorts that had access to a pre-primary in their community, zero otherwise. The unit of observation is community-cohort. Each coefficient corresponds to an individual OLS regression. Trimming and propensity score re-weighting are applied to all regressions (described in section 4.2), including cohort and community fixed effects. Specification (2) includes department-linear time trends, specification (3) municipality-linear time trends and specification (4) time-varying controls (interpolated from the census). Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 3: Impacts of attending a pre-primary, ToT estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Panel A - 2SLS estimates. Dependent variable: % Enrolled in primary at age 12** | | | | |
| % Attended pre-primary | 0.048 | 0.044 | 0.041 | 0.042 |
| | [0.034] | [0.034] | [0.036] | [0.034] |
| R-Squared | 0.024 | 0.027 | 0.062 | 0.029 |
| Number of observations | 8,440 | 8,440 | 8,440 | 8,440 |
| **Panel B - 2SLS estimates. Dependent variable: % Progressing adequately at age 12** | | | | |
| % Attended pre-primary | 0.043 | 0.042 | 0.034 | 0.042 |
| | [0.011]*** | [0.011]*** | [0.011]*** | [0.011]*** |
| R-Squared | 0.065 | 0.073 | 0.106 | 0.073 |
| Number of observations | 8,440 | 8,440 | 8,440 | 8,440 |
| **Panel C - First stage. Dependent variable: % Attended pre-primary** | | | | |
| Pre-primary access (=1) | 0.614 | 0.616 | 0.604 | 0.616 |
| | [0.019]*** | [0.019]*** | [0.019]*** | [0.019]*** |
| R-Squared | 0.627 | 0.631 | 0.657 | 0.631 |
| Number of observations | 8,440 | 8,440 | 8,440 | 8,440 |
| Number of communities | 960 | 960 | 960 | 960 |
| Department-linear time trends | | Yes | | Yes |
| Municipality-linear time trends | | | Yes | |
| Time-varying controls | | | | Yes |

*Notes*: Panel A and B present 2SLS estimates, where the endogenous variable (% Attended pre-primary) is instrumented by the dummy variable Pre-primary access (equals 1 for cohorts that had access to a pre-primary, zero otherwise). Panel C presents the first stage estimates. The unit of observation is community-cohort. Each coefficient corresponds to an individual regression. Trimming and propensity score re-weighting are applied to all regressions (described in section 4.2), including cohort and community fixed effects. Specification (2) includes department-linear time trends, specification (3) municipality-linear time trends and specification (4) time-varying controls (interpolated from the census). Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 4: Heterogeneous impacts of pre-primary access by selected characteristics

| | % Enrolled in primary at age 12 | | % Progressing adequately at age 12 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A - By gender** | | | | |
| Pre-primary access (=1) | 0.043 | 0.047 | 0.025 | 0.025 |
| | [0.028] | [0.028]* | [0.009]*** | [0.009]*** |
| Pre-primary access * Female (=1) | -0.022 | -0.022 | 0.000 | 0.000 |
| | [0.028] | [0.028] | [0.010] | [0.010] |
| **Panel B - By school size** | | | | |
| Pre-primary access (=1) | 0.000 | 0.009 | 0.034 | 0.036 |
| | [0.029] | [0.029] | [0.010]*** | [0.010]*** |
| Pre-primary access * Large school (=1) | 0.058 | 0.054 | -0.018 | -0.021 |
| | [0.036] | [0.037] | [0.011] | [0.011]* |
| **Panel C - By indigenous population** | | | | |
| Pre-primary access (=1) | 0.050 | 0.057 | 0.032 | 0.031 |
| | [0.028]* | [0.028]** | [0.010]*** | [0.010]*** |
| Pre-primary access * High indigenous (=1) | -0.038 | -0.039 | -0.016 | -0.013 |
| | [0.036] | [0.036] | [0.011] | [0.012] |
| **Panel D - By adult population with no education** | | | | |
| Pre-primary access (=1) | 0.051 | 0.062 | 0.036 | 0.035 |
| | [0.025]** | [0.025]** | [0.009]*** | [0.009]*** |
| Pre-primary access * High no education (=1) | -0.046 | -0.055 | -0.025 | -0.022 |
| | [0.036] | [0.036] | [0.011]** | [0.011]** |
| **Panel E - By prevalence of chronic malnutrition** | | | | |
| Pre-primary access (=1) | 0.055 | 0.061 | 0.028 | 0.030 |
| | [0.029]* | [0.029]** | [0.010]*** | [0.010]*** |
| Pre-primary access * High malnutrition (=1) | -0.045 | -0.044 | -0.007 | -0.010 |
| | [0.036] | [0.037] | [0.011] | [0.012] |
| Number of communities | 960 | 960 | 960 | 960 |
| Department-linear time trends | | Yes | | Yes |

*Notes*: Each column in a panel corresponds to a separate regression. In each OLS regression, the relevant explanatory variable, Pre-primary access, is interacted with a dummy for the analyzed characteristic (e.g. in Panel A, girls). In Panels B to E, communities are classified by the median value of the analyzed dimension into high vs. low of the given characteristics. Trimming and propensity score re-weighting are applied to all regressions (described in section 4.2), including cohort and community fixed effects. Even numbered columns include department-linear time trends. Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 5: Take-up heterogeneity of attending a pre-primary

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Panel A - Dependent variable: % Enrolled in primary at age 12** | | | | |
| Pre-primary access * Low take-up | -0.016 | -0.023 | -0.025 | -0.028 |
| | [0.031] | [0.031] | [0.030] | [0.031] |
| Pre-primary access * Medium take-up | 0.038 | 0.037 | 0.042 | 0.040 |
| | [0.036] | [0.035] | [0.033] | [0.035] |
| Pre-primary access * High take-up | 0.060 | 0.060 | 0.058 | 0.059 |
| | [0.031]** | [0.031]* | [0.033]* | [0.031]* |
| R-Squared | 0.023 | 0.026 | 0.061 | 0.028 |
| Number of observations | 8,543 | 8,543 | 8,543 | 8,543 |
| **Panel B - Dependent variable: % Progressing adequately at age 12** | | | | |
| Pre-primary access * Low take-up | 0.003 | 0.001 | -0.001 | 0.001 |
| | [0.009] | [0.009] | [0.010] | [0.010] |
| Pre-primary access * Medium take-up | 0.023 | 0.021 | 0.020 | 0.021 |
| | [0.010]** | [0.010]** | [0.010]* | [0.010]** |
| Pre-primary access * High take-up | 0.045 | 0.046 | 0.039 | 0.047 |
| | [0.010]*** | [0.010]*** | [0.011]*** | [0.010]*** |
| R-Squared | 0.063 | 0.070 | 0.104 | 0.071 |
| Number of observations | 8,543 | 8,543 | 8,543 | 8,543 |
| Number of communities | 960 | 960 | 960 | 960 |
| Department-linear time trends | | Yes | | Yes |
| Municipality-linear time trends | | | Yes | |
| Time-varying controls | | | | Yes |

*Notes:* Pre-primary access is interacted with dummies for groups defined by average community pre-primary take-up rate. The groups (low/medium/high) were classified by terciles. The unit of observation is community-cohort. Each coefficient corresponds to one OLS regression. Trimming and propensity score re-weighting are applied to all regressions (described in section 4.2), including cohort and community fixed effects. Specification (2) includes department-linear time trends, specification (3) municipality-linear time trends and specification (4) time-varying controls (interpolated from the census). Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 6: Robustness checks

| | Baseline estimates | | Only beneficiary communities | | All pre-primary openings | | Communities with average enrollment below 1 | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Panel A - Dependent variable: % Enrolled in primary at age 12** | | | | | | | | |
| *OLS (ITT):* | | | | | | | | |
| Pre-primary access (=1) | 0.030 | 0.028 | -0.001 | -0.005 | 0.010 | 0.010 | 0.019 | 0.013 |
| | [0.021] | [0.021] | [0.024] | [0.024] | [0.016] | [0.016] | [0.024] | [0.024] |
| *2SLS (TOT):* | | | | | | | | |
| % Attended pre-primary | 0.049 | 0.044 | -0.002 | -0.007 | 0.022 | 0.015 | 0.043 | 0.032 |
| | [0.033] | [0.033] | [0.041] | [0.041] | [0.029] | [0.029] | [0.049] | [0.049] |
| Number of observations | 8,543 | 8,543 | 4,620 | 4,620 | 12,435 | 12,435 | 4,014 | 4,014 |
| **Panel B - Dependent variable: % Progressing adequately at age 12** | | | | | | | | |
| *OLS (ITT):* | | | | | | | | |
| Pre-primary access (=1) | 0.024 | 0.024 | 0.016 | 0.013 | 0.022 | 0.018 | 0.021 | 0.019 |
| | [0.006]*** | [0.006]*** | [0.008]* | [0.008] | [0.005]*** | [0.005]*** | [0.007]*** | [0.007]*** |
| *2SLS (TOT):* | | | | | | | | |
| % Attended pre-primary | 0.046 | 0.045 | 0.032 | 0.027 | 0.041 | 0.035 | 0.046 | 0.040 |
| | [0.011]*** | [0.010]*** | [0.014]** | [0.014]** | [0.009]*** | [0.009]*** | [0.016]*** | [0.016]** |
| Number of observations | 8,543 | 8,543 | 4,620 | 4,620 | 12,435 | 12,435 | 4,014 | 4,014 |
| Number of communities | 960 | 960 | 518 | 518 | 1,398 | 1,398 | 453 | 453 |
| Department-linear time trends | | Yes | | Yes | | Yes | | Yes |

*Notes*: Each coefficient corresponds to a separate regression. Both OLS (ITT) and 2SLS (TOT) estimates are presented in each panel. Baseline estimates correspond to those presented in columns 1 and 2 in Tables 2 and 3. Columns 3 and 4 restrict the sample to only beneficiary communities (those that receive a pre-primary). Columns 5 and 6 expand the sample to include all preschool openings during the period 1993 to 2005. Columns 7 and 8 restrict to communities with pre-construction average enrollment below 1. Trimming and propensity score re-weighting (described in section 4.2) are applied to all regressions (except those in columns 3 and 4), including cohort and community fixed effects. Even-numbered columns include department-linear time trends. Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 7: Impacts of pre-primary access by linear distance to nearest community

| Distance to nearest community: | Below median (1) | Above median (2) |
|---|---|---|
| **Panel A - Dependent variable: % Enrolled in primary at age 12** | | |
| *OLS (ITT):* | | |
| Pre-primary access (=1) | -0.001 | 0.003 |
| | [0.025] | [0.025] |
| *2SLS (TOT):* | | |
| % Attended pre-primary | 0.006 | -0.001 |
| | [0.042] | [0.042] |
| Number of observations | 5,537 | 5,572 |
| **Panel B - Dependent variable: % Progressing adequately at age 12** | | |
| *OLS (ITT):* | | |
| Pre-primary access (=1) | 0.017 | 0.022 |
| | [0.008]** | [0.008]*** |
| *2SLS (TOT):* | | |
| % Attended pre-primary | 0.028 | 0.040 |
| | [0.014]** | [0.013]*** |
| Number of observations | 5,537 | 5,572 |
| Number of communities | 626 | 626 |
| Department-linear time trends | Yes | Yes |

*Notes*: Each coefficient corresponds to a separate regression. Both OLS (ITT) and 2SLS (TOT) estimates are presented in each panel. Column 1 restricts the sample to communities below the median linear distance to the nearest community (from 0 to 0.75 kilometers). Column 2 restricts the sample to communities above the median linear distance to the nearest community. Trimming and propensity score re-weighting (described in section 4.2) are applied to all regressions, including cohort and community fixed effects. All estimates include department-linear time trends. Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 8: Falsification test, impacts of pre-primary access before construction

| | Pre-construction cohorts only | | All cohorts: Placebo 1-cohort prior | | All cohorts: Placebo 2-cohorts prior | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A - Dependent variable: % Enrolled in primary at age 12** | | | | | | |
| Placebo (=1) | 0.012 | 0.013 | 0.013 | 0.013 | 0.015 | 0.015 |
| | [0.035] | [0.035] | [0.023] | [0.023] | [0.023] | [0.023] |
| Pre-primary access (=1) | | | 0.037 | 0.035 | 0.041 | 0.039 |
| | | | [0.026] | [0.026] | [0.027] | [0.027] |
| R-Squared | 0.020 | 0.029 | 0.022 | 0.025 | 0.022 | 0.025 |
| Number of observations | 3,778 | 3,778 | 8,543 | 8,543 | 8,543 | 8,543 |
| **Panel B - Dependent variable: % Progressing adequately at age 12** | | | | | | |
| Placebo (=1) | 0.005 | 0.004 | 0.005 | 0.005 | 0.004 | 0.004 |
| | [0.009] | [0.010] | [0.007] | [0.007] | [0.006] | [0.006] |
| Pre-primary access (=1) | | | 0.027 | 0.027 | 0.027 | 0.027 |
| | | | [0.008]*** | [0.008]*** | [0.008]*** | [0.008]*** |
| R-Squared | 0.008 | 0.014 | 0.061 | 0.068 | 0.061 | 0.068 |
| Number of observations | 3,778 | 3,778 | 8,543 | 8,543 | 8,543 | 8,543 |
| Number of communities | 960 | 960 | 960 | 960 | 960 | 960 |
| Department-linear time trends | | Yes | | Yes | | Yes |

*Notes:* In columns 1 and 2 the sample is restricted to the pre-construction cohorts (1988 through 1991) and the variable Placebo is a dummy variable that equals 1 for the 1990/91 cohorts in beneficiary communities. Columns 3 to 6 include the entire sample of cohorts. In addition to the Pre-primary access variable, columns 3 and 6 include a Placebo dummy that equals 1 for the cohort preceding the first exposed cohort and 0 otherwise. In columns 5-6, Placebo equals 1 for the 2 cohorts preceding the first exposed cohort and 0 otherwise. Trimming and propensity score re-weighting are applied to all regressions (described in section 4.2), including cohort and community fixed effects. Even-numbered columns include department-linear time trends. Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

## Table 9: Testing for differential trends in community characteristics

| Dependent variable | (1) | (2) | (3) |
|---|---|---|---|
| % Indigenous | -0.011 | -0.014 | -0.025 |
| | [0.021] | [0.020] | [0.021] |
| % No education | -0.003 | -0.005 | 0.0003 |
| | [0.008] | [0.007] | [0.007] |
| Average age | 0.005 | -0.024 | 0.084 |
| | [0.125] | [0.122] | [0.128] |
| % Household has kitchen | 0.008 | 0.014 | 0.052 |
| | [0.023] | [0.020] | [0.021]** |
| % Household has running water | 0.034 | 0.033 | 0.024 |
| | [0.024] | [0.024] | [0.028] |
| % Household has concrete floor | -0.016 | -0.017 | -0.007 |
| | [0.010] | [0.010]* | [0.010] |
| Community size | 0.770 | 3.770 | 26.324 |
| | [16.551] | [16.214] | [18.738] |
| | | | |
| Dummies | None | Department | Municipality |
| Number of communities | 960 | 960 | 960 |

*Notes:* Each coefficient corresponds to a separate OLS regression. Estimated coefficients correspond to regressing the change between 1994 and 2002 in a given community characteristic (e.g. % indigenous) on a dummy variable that indicates whether the community had access to a pre-primary. Column 2 includes department dummies while column 3 includes municipality dummies. Trimming and propensity score re-weighting are applied to all regressions (described in section 4.2). Standard errors, reported in brackets, are clustered at the community level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.