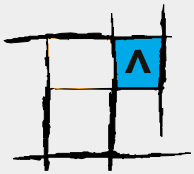
LACEA 2007
MONTEVIDEO URUGUAY


Luis Rubalcava, joint with Susan Parker and Graciela Tervel

# Educational Attainment of Indigenous Children in Mexico: A Problem of Family Resources or Language Barriers?



*Preliminary and Incomplete*

# Educational attainment of indigenous children in Mexico: A problem of family resources or language barriers?

Susan W. Parker, CIDE

Luis Rubalcava, CIDE

Graciela Teruel, UIA

June, 2001

Revised version: August, 2001

This paper was written with financing from the Inter-American Development Bank from the Network Center Project on Social Exclusion. We thank Jere Behrman, Alejandro Gaviria and Miguel Szekeley for useful comments and Rodolfo Islas and Allan Pasalagua for outstanding research assistance. The opinions presented here represent the viewpoint of the authors only and do not represent the views of their respective institutions or the IDB.

ABSTRACT:
This paper analyzes the educational attainment of indigenous children in Mexico. Using large household data sets from rural communities where a majority of indigenous people live in Mexico, we analyze the potential explanatory factors for low educational attainment of indigenous children. We find that, overall, indigenous children fare worse than their non-indigenous classmates. Nevertheless, there is important heterogeneity within the indigenous group. In particular, monolingual indigenous children (those who speak only an indigenous language) do much worse in school than bilingual indigenous children who speak Spanish as a second language.

Using community and instrumental variable models which control for the possible endogeneity of languages spoken within the indigenous population, we are able to shed some light on the reasons for this poor performance. While controlling for parental and community variables reduces the overall negative effect of speaking only a native language, these effects remain significant. We interpret these results as evidence that while family resources and school quality are clearly important, they cannot explain all of the differences in educational attainment between bilingual indigenous and monolingual indigenous children. Rather, language barriers represent an important aspect of barriers that indigenous children face in school.

In order to better understand the extent to which these language barriers affect indigenous children's schooling outcomes we further examine the possible role of bilingual education in improving the educational performance of indigenous children. Our initial results demonstrate that indigenous primary schools in Mexico, which practice bilingual education, improve the educational performance of monolingual children at the primary level although the effects are thus far relatively small in magnitude.

## Introduction

Over the years Mexico has experienced important advances in its social indicators. At the beginning of the twentieth century the literacy rate was only 22 percent. [INEGI (1994)]. Now, however, almost 9 Mexicans out of 10 are literate [Scott (2000)]. Likewise, whereas in 1910 the average years of completed schooling was 2.8, in 1990 it had increased to almost 7 years. [Barro & Lee (1996), and Scott (2000)]. These changes have been accompanied by rapid population growth and a heterogeneous demographic regional composition. Today, 75 percent of Mexicans live in urban areas, whereas only 25 percent live in remote rural communities. [INEGI (2000)].

These general increases in human capital formation have, nevertheless, been associated with limited reductions in income inequality[1]. Social progress has been far from homogeneous. Whereas urban areas have to a large extent seen improvements in their social indicators, remote rural communities, with consequently high costs of bringing basic public infrastructure to their inhabitants have continued to lag far behind urban areas. These isolated communities are the home of the vast majority of indigenous groups.

In 1994, the deprived economic conditions of indigenous people, led to a social movement against the *status quo* in Chiapas, a southern, highly indigenous, state in Mexico. Since then, more public infrastructure –in terms of basic services, health centers and schools – has flown to these marginal communities in an effort to reduce the poverty of their inhabitants. Nevertheless, the potential problems and limitations which indigenous children may face in school is a subject on which little evidence exists in Mexico. Assuming that, in fact, the education attainment of indigenous children is lower than non-indigenous children, there are a large number of potential explanatory factors.

[1] In 1950 the country's Gini coefficient was 52 but by 1990 it had increased to 60.5. [Scott, (2000)].

These include lower family resources, access to lower quality schools, discrimination as well as cultural and language barriers. *(Caso et al, 1981).*

The current paper analyzes to what extent being indigenous is associated with lower schooling outcomes for children living in remote rural communities. We compare different schooling indicators among children with the same socioeconomic opportunities, household demographics and community characteristics, age and gender, but with different ethnic backgrounds. For the analysis, we exploit unique household information on households in rural Mexico which was carried out in poor rural communities. These communities contain a majority of indigenous households in Mexico. We also use a nationally representative household survey to carry out the same analysis to insure that our results are comparable.

To our knowledge, this is one of the first studies to specifically analyze the determinants of indigenous educational outcomes. Whereas there is some previous descriptive evidence suggesting that indigenous children tend to have lower educational outcomes than non-indigenous children, the potential factors associated with this low performance have not been studied. If indigenous households are poorer than other households and poverty is a factor affecting school decisions, then simple correlations cannot determine whether the poor performance among indigenous children is due to low family resources or other causes, which may include cultural factors, language barriers or access to lower quality schools. In this paper we are able to shed some light on the extent to which family resources versus language barriers are related to the lower educational outcomes of indigenous children.

Our results show that indigenous children do indeed fare worse than their non-indigenous classmates even within the relatively homogenous rural marginated communities of our sample. Nevertheless, there is important heterogeneity within the indigenous group. In particular, monolingual indigenous children (those who speak only an indigenous language) do much worse in school than bilingual indigenous children who speak Spanish as a second language.

To shed some light on the reasons for this poor performance, we carry out regression models of the determinants of children's schooling outcomes. We first control only for background family characteristics and resources at the household level, followed by community effects. We then use instrumental variable methods to explore the possible endogeneity of language spoken within the indigenous population. We find that, while controlling for parental and community variables reduces the overall size of the negative effect of speaking only a native language, these effects remain significant. We interpret these results as evidence that while family resources and school quality are clearly important, they cannot explain all of the differences in educational attainment between bilingual indigenous and monolingual indigenous children. Rather, language barriers represent an important aspect of problems that indigenous children face in school.

In order to better understand the extent to which these language barriers affect indigenous children schooling outcomes, we further examine the possible role of bilingual education in improving the educational performance of indigenous children. The Secretary of Public Education offers indigenous primary schools which include bilingual teachers as well as textbooks in native languages. We study whether the availability of indigenous schools increases the likelihood of indigenous children attending school and whether, to some degree, it compensates or reduces the language barriers described above. Using community fixed effects estimators which control for possible biases due to endogenous program placement, our results suggest that the educational disadvantage due to language barriers is reduced for children who have the option of attending a primary indigenous school.

These results have important policy implications for indigenous learning in Mexico. We have shown that controlling for an important number of measures of family resources, access to schools and community characteristics, indigenous monolingual children continue to perform worse than their bilingual counterparts as well as non-indigenous children. In addition, we demonstrate that indigenous schools, which practice

bilingual education, improve the educational performance of monolingual children at the primary level. Thus, while higher poverty levels contribute to differences between indigenous and non-indigenous educational outcomes, they do not explain all of the differences. The analysis of this paper suggests that a large fraction of this worse performance is due to lack of knowledge of Spanish. The policy prescriptions would thus call for study of the best ways to promote learning of indigenous children, and in particular the learning of Spanish. To the extent that indigenous primary schools seem to improve the performance of indigenous children, expansion of these integrated educational programs would seem to be warranted. One possible caveat for the future is that their effectiveness may be reduced if indigenous primary schools decrease the probability that indigenous children learn Spanish.

## 2. Background

To our knowledge, there has been little previous research in Mexico on the educational attainment of indigenous children. One exception is Panagides, 1999 who uses the Survey of Income and Expenditures (ENIGH) to look at various economic dimensions of indigenous individuals and families, including educational attainment and earnings. Nevertheless, since this survey contains no information on whether individuals are indigenous, the indicator constructed to measure indigeneity is a community based indicator, defined by the overall percentage of individuals speaking an indigenous language in the municipality of residence.

Lopez (1999) has analyzed the impact of the PARE program (*Programa para abatir el Rezago Educativo*), which gave additional school resources -- such as textbooks, and teacher training -- to schools in Mexican states with high rates of poverty and low educational attainment on student test scores in math and Spanish at the primary level. Prior to the program, test scores were lowest for children enrolled in "indigenous" schools. As a result of the program, improvements in test scores were shown in all areas, with the greatest improvements occurring in indigenous schools, although even after the program, test scores in indigenous schools remained lower than in other primary schools

in rural areas. While insightful, the study does not have the richness of the household level data that we use to study this topic.

The meaning of indigenous is a complex subject, involving cultural traditions, languages and practices which have developed over centuries. In the case of Mexico, there are 62 different ethnic indigenous groups, speaking over 80 different languages and with different sets of traditions. While ideally, our definition of indigenous would be multi-dimensional, including not just language but other indicators as well, due to data constraints we are restricted to definitions based on language spoken.

We are fortunate, nevertheless to have data which includes individual definitions of language spoken. Each individual is asked if they speak an indigenous language. Those that report they do speak an indigenous language are then asked if they also speak Spanish. In this way, we can make the important distinction between indigenous children who speak only a native language versus indigenous children who are bilingual.[2]

We are also fortunate to have separate indicators of whether the parents of an indigenous child speak only an indigenous language or are bilingual. While parental language is highly correlated with the language spoken by the child, there is some important variation and in particular between the father and mother. It is much more common for fathers to be bilingual than mothers.

In this paper we exploit the richness that individual level data offers to analyze the impact of the condition of being indigenous on schooling outcomes of children. By making use of the variation that exists between mother, father and child languages spoken, we are able to control for endogeneity of languages spoken within the indigenous population. That is, we are able to take into account the fact that learning Spanish may be a choice.

[2] Ideally one would prefer a more objective measure of indigenous status—one not dependent on self-reporting, which is potentially susceptible to a stigma effect. If indigenous do not accurately report their real status, biases may result in the estimation of the differential schooling outcomes between non-indigenous and indigenous children.

## 3. Descriptive Analysis

We begin with an overall description of the indigenous population in Mexico. Using a nationally representative sample of the Mexican population, approximately 5 percent of all children are indigenous and of these, 70 percent report speaking Spanish. The national survey also shows that a majority of the indigenous population are located in rural areas. Within urban areas, only 1.1 percent of children are indigenous versus 11.8 percent of children in rural areas.[3] (See Tables 1A and 2A in the Appendix.)

Using non parametric regression[4] with a national sample of the Mexican population, Figure 1 illustrates the relationship between completed years of schooling for three groups of children: non-indigenous, indigenous monolingual and indigenous bilingual. The graph shows little differences among the three groups at ages below 8, and larger and increasing differences afterward. As expected, non-indigenous children show the highest achievement of the three groups, followed by the bilingual group. The indigenous monolinguals, however, lag behind at all age groups. In general, this graph indicates that indigenous children who remain monolingual achieve very low levels of education on average, while indigenous children who learn Spanish over time (bilingual), perform better, although not as well as non-indigenous children. This is true across the whole child-age distribution.

---

[3] It should be noted that the national sample of the ENCASEH only contains information on 9910 households so that the number of indigenous cases is quite small. The number of indigenous children between the ages of 6 and 18 is 658 and less than 200 of these report speaking Spanish. Thus, the nationwide percentages of the indigenous population may differ compared with other nationally representative surveys, such as the Census. We use the nationally representative ENCASEH only to insure that our results are not overly biased by focusing on a sample which is not representative at the national level.

[4] The non-parametric estimator we apply carries out locally weighted, smoothed scatter plots (LOWESS). In this procedure the regression is weighted so that the point in the middle gets the highest weight and points farther away receive less weight. This local average depends on the amount of smoothing, which in turn is affected by the choice of bandwidth *h*, as in

$$f_k = \frac{1}{nh}\sum_{i=1}^{n} K\left[\frac{x - X_i}{h}\right]$$

where K was chosen to be the Epanechnikov since it has the property that it is most efficient in minimizing the mean integrated squared error.

Figure 2 shows the same graphs as in figure 1 but for a sample representative of only rural areas, which is where most of the indigenous population is concentrated. Interestingly, the graph shows little difference in terms of years of completed schooling between non-indigenous children and bilingual indigenous children. There are, however, huge differences between indigenous monolingual children and the other two groups above the age of 8. By the age of 18, the average indigenous monolingual child has achieved only about 2.5 years of completed schooling versus the other two groups which achieve more than double the level, on average about 7 years of schooling. The results from this regression foreshadow our regression analysis, which will demonstrate the large importance of language in determining educational outcomes.

We now illustrate the extent to which learning Spanish for the indigenous population is likely a dynamic phenomenon. Using non-parametric analysis on the probability of being monolingual for indigenous children, Figure 3 shows that indigenous children who lag behind in school are those who are unlikely to learn a second language. This implies that bilingual is a dynamic concept and integrally related with school attendance. As children participate in school for a given age, there is obviously learning occurring in terms of languages. Our sample also shows that whereas 37 percent of indigenous children at the age of 6 speak only an indigenous language, by the age of 18, only 10 percent of indigenous children are monolingual. Our econometric analysis will thus treat the learning of Spanish by indigenous children as endogenous to schooling outcomes.

## 4. Data Description

The data to be used for this project comes directly from data carried out through the Mexican Education, Health, and Nutrition Program (PROGRESA). Progresa is a large anti-poverty program in Mexico, implemented in poor rural areas and providing monetary and in-kind benefits linked to regular school attendance of children and health clinic visits of the family. The program has collected a great quantity of socio-economic information as a result of both its mechanism of selection of beneficiaries and its evaluation. We use two principal, related sources of information for the analysis, which we now describe.

The targeting mechanism of Progresa involves carrying out a socio-economic questionnaire (Survey of Household Socio-economic Characteristics-ENCASEH) for all rural isolated communities eligible for Progresa. There are currently about 3 million ENCASEH surveys which include information on educational attainment, monetary income, durable goods, labor force participation as well as indigenous status. Our first data source is a random sample of this survey (equal to about 120,000 households) which provides cross-sectional information for all 32 Mexican states. While only cross-sectional, this survey has the advantage of providing a vision of the indigenous population living in marginalized rural areas in all of Mexico. While it is not representative at the national level, the survey does capture a majority of all indigenous households in Mexico, we estimate approximately 60 percent.

Nevertheless, to insure that our results are valid in making inferences about the Mexican population -- perhaps due to possible sample selection (for instance, if indigenous individuals in marginated areas are not representative of all indigenous individuals) -- we also use a nationally representative survey. We are fortunate that a separate national sample of the same ENCASEH questionnaire was carried out in 1997

and provides a convenient way to compare our results. This national sample includes 9,910 households and is representative of both urban and rural areas in Mexico. [5]

Our dependent variables address short and long term educational outcomes for boys and girls between the ages of 6 and 18: 1) enrollment and 2) years of completed schooling, respectively. [6]

Finally, we supplement our household and student level data with school level information from the Secretary of Public Education (SEP) which allow us to link the characteristics of available schools to children's educational outcomes. This data comes from a census collected by the SEP and contains specific information about each school, such as number of pupils, grade averages, education of teachers and characteristics of the school infrastructure. In particular, at the primary level, we take advantage of information on indigenous primary schools to analyze the effect of bilingual education on the educational achievement of indigenous children.

Using our 32 state sample of the ENCASEH, which has 127,844 families, 29.2 percent (37,346) of the heads of these households report speaking an indigenous language. Of these household heads, 87.7 percent (32,435) also report that they speak Spanish, suggesting that a minority of household heads speak *only* an indigenous language. Of children aged 6 to 18, 23.8 percent report speaking an indigenous language and of these, 81.7 percent also report speaking Spanish. It is clear, thus, that this 32 state sample has a much larger concentration of indigenous families than at the national level.

We close this section with a comment on the distribution of the indigenous population. As one might expect, most rural communities tend to be primarily

[5] To avoid confusion, this survey we will refer to as the national ENCASEH whereas the ENCASEH drawn from the rural communities we will call the 32 States ENCASEH.

[6] We also performed analysis using the schooling gap measure, defined as age-years of schooling –6 and is an indicator of the extent to which a child is "behind" where he/she should be in school. The lack of presence of non-linearities in the relation between age and education resulted in schooling gap estimates quite similar to those obtained with years of schooling, so we do not report the former. These results showed that indigenous children have, on average, a schooling gap 2.3 percent higher than non-indigenous children.

indigenous or primarily non-indigenous. Whereas our 32 state sample does not include census of entire communities, we are able to gain some idea of the extent of isolation or segregation through decompositions of the variance of the variable indigenous (defined as speaking a indigenous language). Our analysis shows that about 86% of the variation with respect to indigenous derives from variation between communities. This will have implications for the use of community fixed effects models below which rely only on the variation within communities. Of the 26,079 communities with at least one child in the sample, only about 10% of these communities have both indigenous and non-indigenous children in the sample.

## 5. Methodology and Results

*Disentangling economic conditions from other factors*

Our previous descriptive non-parametric results clearly showed that indigenous monolingual children lagged behind in schooling outcomes. However, this descriptive evidence cannot distinguish between whether the poor school performance of indigenous children simply reflects cultural and language barriers, or whether it simply reflects the likely inferior social and economic factors which indigenous households face. Disentangling these effects is crucial for policy making: if the poor school performance of indigenous children is mainly driven by the poor economic conditions in which they live, then anti-poverty programs would be largely sufficient to reduce the education gap between indigenous and non-indigenous children. However, if the poor schooling outcomes are the result of other structural factors, such as a language or cultural barrier – holding poverty levels constant- then social programs aiming only at improving the marginality conditions of indigenous communities will not be sufficient.

In order to begin to disentangle the effect of a language barrier from that of socioeconomic resources, we now turn to a regression analysis. We start with assessing the association between schooling opportunities — as measured by enrollment and years of completed schooling -- and belonging to an indigenous group. We estimate the following relationship for each household child in our sample:

$$(1)\ S_{ic} = \beta_0 + X_{1ic}\beta_1 + X_{2ic}\beta_2 + \delta_1 INDIG_{ic} + u_c + e_{ic}$$

Where $S_{ic}$ stands for the education outcome variables of the child (i) in community *c*, $X_1$ represents his/her observed characteristics including age and sex ; $X_2$ represents a set of household characteristics including mother and father education, age; measures of household wealth and dwelling characteristics. Measures of household wealth and dwelling characteristics include: ownership of land, access to water and electricity, whether the floor of the house is made of cement and ownership of durable goods such as refrigerator and stove; INDIG is an indicator of whether the child belongs

to an indigenous group. The model also includes a community fixed effects $u_{c,}$ given that failing to control for observed and unobserved time-invariant community characteristics which may be correlated with indigenous child schooling outcomes, such as local infrastructure, market prices, cultural community behavior and overall economic conditions, could bias our results. $e_{ic}$ corresponds to an error component that reflects all remaining unobserved characteristics of the model.

The particular hypothesis we are interested in testing relates to the existence of any form of social exclusion with respect to school productivity of children that belong to an indigenous group, holding everything else constant. That is, we test whether $d$ is different to zero. A negative coefficient would imply a negative effect of group membership with respect to school opportunities, thus suggesting indigenous children are in a disadvantaged position relative to their non-indigenous classmates.

Table 1 shows the determinants of years of schooling and in particular, the effect of being indigenous. Here, we begin with a general measure of indigenous which includes indigenous children who are either bilingual or monolingual. In the analysis below, we will separate the two groups.

Table 1 reports a number of specifications, beginning with a minimal specification in which years of completed schooling only depends on child characteristics. We then progressively include parental characteristics, household wealth indicators and community effects. This allows us to analyze the extent to which the impact of indigeneity is altered by separately including these characteristics of the household. Column (1) shows that an indigenous child lies on average half a year behind relative to his/her non-indigenous classmates with the same age. As expected, years of completed schooling is a monotonic function of the years of age of the child. Column (2) adds the age of the parents in years. Children with older mothers tend to be more educated, but only marginally. In order to see--whether the maternal age effect is contaminated with a human capital effect -older parents tend to be less educated since they belong to older generations--we further control, in column (4), for parental levels of

education. The effect of mother's age rises marginally and the father age effect becomes positive and significant. As expected, children whose parents have higher levels of education are more likely to have higher years of completed schooling. This may reflect parental ability in child rearing or economic conditions of the household, as parental human capital is also a measure of permanent income. It is worth noting that the magnitude of the coefficient of indigeneity decreases, suggesting that the effect of being indigenous is highly correlated to household resources.

Column (5), along with parental characteristics, further controls for household wealth. In particular, we include controls for whether the dwelling has concrete floors, walls and ceilings, whether the household has access to running water and electricity, and of whether the household owns agricultural land. All wealth measures are significant and have the expected sign. However and perhaps more importantly, the comparison of columns (5) and (1) show that the impact of being indigenous on educational outcomes has been reduced by half. That is, controlling for parental and household characteristics demonstrates that at least half of observed differences between indigenous and non-indigenous is primarily due to family background and other economic variables.

So far, we have not taken into account community characteristics, such as schooling quality. It is likely that indigenous households live in poorer communities with an inferior schooling infrastructure than the rest of the population. Therefore, failing to control for community observed and unobserved heterogeneity could cause an overestimation the indigenous impact as well as the effect of household resources on child schooling outcomes. To correct this problem, column (6) introduces community fixed effects. It is interesting to note that when interpreting the impact of being indigenous, apparently household resources capture most of the community effects, given that the indigenous coefficient does not change when community controls are added.[7]

[7] The household wealth coefficients change in magnitude but not in sign. Under community fixed effects, the coefficients on water and electricity and concrete dwelling characteristics are reduced, contrasting with the increase in the effect of owning agricultural land. The differential change of the wealth variables may be related to the fact that wealthier households with concrete dwelling and public services availability are

Column (7) presents the same specification as that in column (6), but now we use our *national* sample, restricting attention to only rural areas. This resulting sample is representative of all rural areas in Mexico.[8] The estimated effects of indigeneity are about double the size than those reported in our ENCASEH 32 state sample. This is perhaps not surprising as this (national) sample is much more heterogeneous than our sample which restricts analysis only to very marginated communities. However it is also likely that with a more heterogeneous sample, we are less successful at controlling for a household's economic (unobserved) circumstances, and thus part of the estimated effect of being indigenous in this national sample reflects uncontrolled economic factors at the household level. We continue the rest of our analysis with the ENCASEH 32-state sample given that households and communities in it are overall poorer and more homogeneous, which allows us to better control for household resources and thus better isolate the impact of being indigenous.

It is obvious that completed years of schooling for boys and girls between the ages of 6 and 18 is a longer term indicator, as opposed to current school enrollment, which is short term. Table 2 presents specifications (5) and (6) for our second schooling outcome: current school enrollment. For comparability, we also present the last two specifications of Table 1 as the first two columns of Table 2. Columns (3) to (4) correspond to OLS and community fixed effects specifications for the probability of current enrollment, respectively.[9] In general, the results are similar compared to those using years of completed schooling in that controlling for household level variables and community fixed effects, indigenous children remain likely to do worse than their non-indigenous classmates. Column (2) shows that indigenous children are, on average, 3.2

located in richer communities, whereas agricultural oriented households are more likely to live in rural communities with relatively less development.

[8] We exclude urban areas from the national sample because of the very low proportions of indigenous children in urban areas. Of the 8,978 urban children aged 6 to 18 in our ENCASEH-national sample, only 101 report speaking an indigenous language.

[9] For comparability purposes and given our large sample size, we exploit that the Linear Probability Model (LPM) coefficients are consistent, and estimate the probability of school enrollment using OLS models.

percent less likely to be enrolled in school even after controlling for household and community characteristics.

Note that the coefficient of being indigenous on the likelihood of currently attending to school reverses in sign after we control for community fixed effects. This is not the case for years of schooling, suggesting that community unobserved characteristics are differentially correlated to schooling attendance of indigenous children. The last two columns of Table 2 try to clarify these differences. In Columns (3) and (4) we run a different specification where we divide the indigenous into two groups, *monolingual* indigenous and *bilingual* indigenous to show the differential impact on school enrollment.[10] The effect of being monolingual for indigenous children increases in magnitude but remains negative when including community fixed effects. This is not the case for the bilingual indigenous coefficient, which turns from being positive and significant to negative and significant. These results – which are meant only to be suggestive-[11] show that community unobserved characteristics affect differentially the schooling enrollment of monolingual indigenous children and bilingual indigenous children.[12] Therefore, the regressions presented above are, to some extent, an average effect of these two "types" of children and may mask important differences between these two groups. Furthermore, the persistent negative effect of being indigenous monolingual [Table 2: columns (3) & (4)] is suggestive that indigenous children are not only economically disadvantaged, but may also face other cultural or language barriers when attending school.

*Determining schooling outcomes among indigenous children*

[10] As in previous specifications, non-indigenous is the omitted child category.

[11] As discussed further below, we consider the language division to be endogenous to the determinants of educational outcomes.

[12] We have explored in more detail the characteristics of the communities and found that an important community characteristic explaining this reversal is the percentage of indigenous children in the community and in particular the percentage of those speaking only a native language. This community variable is however obviously endogenous to our model and thus we do not present this specification in the regressions.

Next, in an effort to disentangle the possible cultural or language barrier effect from other factors, we restrict our sample to those children who are indigenous, dividing them into two groups: those who speak Spanish (bilingual) and those who are only monolingual. Here, our interest is to examine the extent to which there is heterogeneity among the indigenous population and the extent to which learning or not learning Spanish affects children's performance in school. We use the following specification:

$$(2)\ S_{ic} = \beta_0 + X_{1ic}\beta_1 + X_{2ic}\beta_2 + \delta_2 MONO_{ic} + u_c + \varepsilon_{ic}$$

Where $S_{ic}$, $X_{1ic}$ $C_{1ic}$ and $X_{2ic}$ are as defined above, *MONO* is an indicator of whether the indigenous child is monolingual versus the alternative of speaking Spanish as a second language, $u_c$ is a community fixed effect, and $\varepsilon_{ic}$ corresponds to all remaining unobserved characteristics.

Table 3 presents our findings on the impact of only speaking a native language on years of completed schooling and school enrollment, for indigenous children. Columns (1) through (4) show varying specifications relating to the inclusion and exclusion of speaking only the native language, relative to the excluded category of also speaking Spanish as a second language. Column (1) shows OLS estimates of the impact of a child's language controlling only for child characteristics, whereas column (2) includes parental and household characteristics and community fixed effects as well. Column (3) excludes the language spoken by the child and includes the languages spoken by the parents while column (4) includes the languages spoken by both the child and the parents.

Columns (1) and (2) make clear that the language of the child has very large negative effects on schooling outcomes relative to other indigenous children that do speak Spanish. For example, an indigenous child who does not speak Spanish lies (on average) 1.1 years behind in terms of completed years of schooling relative to his/her bilingual classmates with the same household and community resources [Column 2]; and

he/she is 14 percent less likely to currently be enrolled in school [Column 4]. These results make clear the great educational disadvantage resulting when indigenous children do not learn Spanish relative to his/her indigenous companions who do speak Spanish. It is noteworthy that these differences are much greater than the overall differences previously observed between indigenous and non-indigenous children (Table 2).

Furthermore, specifications (3) and (4) show that the language barrier effect only operates through the child's ability to speak Spanish. Column (3) shows that while parental language has significant impacts on child's educational outcomes, the (absolute) size of the coefficients is much smaller than that when we in addition take into account that the child is monolingual. Moreover, Column (4) demonstrates that the child's monolingual effect is robust to the inclusion of controls of whether parents are also monolingual. Mother's and father's ability to speak the language has no significant effect on the child's human capital assessment after child language is controlled for. These two facts suggest the child's "monolingual" effect is likely to reflect more a language barrier at school, rather than a parental or household (unobserved) cultural factor.[13]

It is important to emphasize here, however, that the variable *MONO* is unlikely to be exogenous to schooling determinants. A child's ability to speak a second language (Spanish) is likely to be highly correlated with children's enrollment and school attendance as well as with previous decisions regarding schooling of the child, that is, it is in some sense a choice variable. [See Figure 3]. Consequently, failing to control for unobserved characteristics *at the household level*, such as parental tastes or parental child-rearing ability in human capital formation, may lead to an overestimation of the true language barrier effect if the decision to send the child to school is correlated with the characteristics of the parents. To overcome this problem, we treat the language spoken by the child as endogenous and instrument the child's probability of being monolingual with his/her parents' ability to speak Spanish as a second language. [14]

[13] This result will be key to identifying the "true" language effect, when using instrumental variable methods.

[14] An interesting, largely theoretical work has developed on motivations and incentives for learning a second language. (See Lazear, 1999, Lang, 1986, Church and King, 1993). Lazear argues that incentives to learn a second language and other forms of assimilation are lower when the large majority of the group speaks the native language. However, Church and King argue that the overall benefits of language acquisition are increasing in the number of individuals who speak the language, giving rise to possible

We argue that mother's and father's ability to speak Spanish is a good instrument for a child's ability to speak Spanish. First, we have shown that mother and father's ability to speak Spanish does not significantly affect their child's schooling outcomes, once we control for a child's own ability to speak Spanish. Second, the child's probability of speaking Spanish is highly correlated to her/his parents´ language ability.[15] Finally our Basmann (1960) IV over-identification tests show that parental language ability is a good instrument to identify the structural model.

Columns (5) and (7) in Table 3 present 2SLS estimates of the child's lack of ability to speak Spanish on completed years of schooling and on the probability of current enrollment in school respectively. With both indicators, the effect remains negative and significant, although decreases in magnitude. For example, 2SLS estimates on the child's probability of school enrollment, show that failing to control for potential endogeneity would overestimate its effect by almost 50 percent (in absolute terms): 2SLS estimates show that a child that does not speak Spanish is "only" 10 percent less likely to attend to school, as opposed to an OLS community fixed effects (negative) probability of 14 percent. [See table 3: columns (7) and (6), respectively]. Moreover, Hausman specification tests reject the exogeneity of the child being monolingual as an explanatory variable.

One final exercise that we carry out is to examine how the language barrier may change or accumulate over time. Our previous descriptive results had suggested a widening difference in years of schooling completed between monolingual children and bilingual as well as non-indigenous children. We thus repeat our last specification in Table 3 above, that of instrumental variables estimations, of the determinants of years of completed schooling and school enrollment by age group. Table 4 presents the summary of these results which show that initially (for the age group 6 to 8) the effect of being

externalities which in turn imply that language acquisition decisions by the population may be inefficient.

[15] Coefficients of determination of the first-stage regressions of the probability of the child being monolingual against parental Spanish knowledge, show $R^2$*s* of 42 percent for years of completed schooling and 23 percent for school attendance, respectively.

monolingual is relatively small for indigenous children on their years of completed schooling, reducing years of schooling in only about one/tenth of one year. The effects of this language barrier, however, increase with age, so that by the age group 15 to 18, the average difference due to language barriers is more than 2 years of schooling. With respect to school enrollment, it is interesting to note that the estimated negative effect of being monolingual is large (about 10%) and apparent even in the early ages (reflecting to some extent a later entry to school of monolingual children relative to bilingual children). This negative effect continues to be strong over the different age groups with the exception of the oldest age group (15 to 18) where the estimated effect of being monolingual is no longer a significant determinant of school enrollment. [16]

In summary, we conclude that failing to control for endogeneity factors leads to an overestimation of the language barrier effect. Nevertheless, this effect remains quite large, and suggests that the language barrier results in a significant disadvantage in terms of overall human capital acquired by indigenous children, a disadvantage which accumulates as children age. Whereas most indigenous children do in fact learn Spanish, those that do not, achieve much lower levels of education, which are likely to have additional long term effects, carrying over into higher levels of extreme poverty when adults. [17]

*Initial evidence on bilingual education*

[16] This does not imply of course that there are no crude differences between school enrollment of monolingual and bilingual children at the ages of 15 to 18, but rather that these differences are no longer due to language barriers or other cultural differences.

[17] Models carried out with gender interactions reveal a number of interesting gender differences which, for reasons of space, we can only briefly comment here. Replicating the results in Table 3 for girls and boys, the estimated negative effect of being monolingual on years of schooling is approximately 20% larger for girls than that of boys. It is interesting to note that while the proportions of indigenous girls who learn Spanish is comparable with that of boys until about the age of 15, thereafter, a lower proportion of girls report learning Spanish than boys. This coincides with an increasing gap between boys in girls in years of completed schooling and school enrollment during these ages. The largest gaps between boys and girls are evident in the monolingual indigenous population, with smaller gaps in the bilingual indigenous population and little or no gender gaps within the non-indigenous population. These patterns of school attainment and language acquisition are also somewhat related with marriage patterns, as by the age of 15, a significant minority of indigenous girls begin to marry.

The previous analysis has made abundantly clear that indigenous children who learn Spanish have much higher achievements in schooling than those who remain monolingual. We have argued this is a strong indicator of language barriers.

To further explore the nature of language barriers and their possible relationship with bilingual education, we exploit a dichotomy that currently exists in the Mexican public educational system, which provides *Spanish-type* and *indigenous-type* school programs in rural communities. The Secretary of Public Education (SEP) currently operates bilingual schools in a number of indigenous communities at the primary and pre-school level. These schools include bilingual teachers with textbooks in the indigenous language, with the goal to "favor the acquisition, strengthening and development of the indigenous languages as well as the Spanish language by avoiding the imposture of one language over the other", [Caso A., Zavala J.M. and M. González, (1981)]. Proponents of indigenous education generally suggest that bilingual schools encourage the educational attainment of indigenous children by reducing their levels of discouragement in class and by promoting bilingual teachers who speak a child's indigenous language to pay more attention and discriminate less against students who do not speak Spanish.

If the large educational difference between monolingual and bilingual children is largely a language effect, we would expect that an indigenous school would reduce the negative effect of language between monolingual and bilingual indigenous children in their educational attainment. To test for this possibility, we estimate the following equation:

$$(3)\quad S_{ic} = \mathrm{B}_0 + X_{1ic}\mathrm{B}_1 + X_{2ic}\mathrm{B}_2 + d_2 MONO_{ic} + d_3 IND\_PRIM_c + d_4 MONO_{ic} * IND\_PRIM_c + u_c + e_{ic}$$

The model is an extension of model (2) which interacts the child being monolingual (as opposed to bilingual), with an indicator variable, *IND_PRIM,* that takes the value of one if the community provides an indigenous primary school.

The coefficient $d_4$ is the coefficient of interest, and is a double difference estimator. It tells us whether the language gap in education between monolingual and bilingual children is different for children with an indigenous primary school in their community as

opposed to those who do not have an indigenous primary school.[18] If our language barrier hypothesis is correct, we would expect this coefficient to be positive. That is, assuming that indigenous schools reduce the language barrier, one should see a narrowing of the educational disadvantage between bilingual and monolingual indigenous children in communities with indigenous schools.

To test specification (3), we merge our ENCASEH-32 states data with data from the Secretary of Public Education (SEP) from 1997 and define whether an indigenous primary school is available for children in the community where they live. Availability is initially defined using the school which is closest (in kilometers) to the community where the child lives. At the primary level, this is normally the school or schools located within the community, as over 80 percent of communities have at least one primary school within their community. When there is no school located within the community, we calculate the distance to the nearest community with a school in kilometers -with a maximum distance of up to 5 km-[19] and we use the characteristics of this (or these) schools to represent the available supply of indigenous schools for the child.

In the hopes of better capturing behavior in terms of the school that children actually attend, our empirical model restricts attention to communities where there is a primary school located within, and not outside, the community under the above criteria, and we also constrain our sample to communities where there is only one available primary school. According to this procedure, we find that 55 percent of indigenous children in our sample have access to an indigenous school. In distinguishing by whether the child is bilingual or not, we find that 51.8 percent of bilingual children have access to an indigenous school whereas the figures for monolingual indigenous children rises to71.3 percent.

Our models thus tests the language barrier hypothesis on the determinants of years of schooling and enrollment of children eligible for primary school only (children with

[18] $d_4 \equiv \{ \left[S_{monolingual} - S_{bilingual}\right]^{Indi.School} - \left[S_{monolingual} - S_{bilingual}\right]^{Spanish.School} \}$

[19] This is done through the use of Geographical Information Systems (GIS) software.

less than six years of completed schooling).[20] Given the short time in operation of indigenous schools, it is likely that if any effects are to be found, they would be found on our short term schooling indicator: enrollment. Table 5 summarizes our findings for both completed years of schooling and current school enrollment. Consistent with the evidence presented in table 3, monolingual indigenous children attain poorer schooling outcomes than bilingual children, independently of the educational system.[21] At first glance, OLS results on years of completed education [Column (1)] suggest that there is no difference in the *magnitude* of the learning gap between children attending indigenous schools and those attending Spanish schools.

Nevertheless, school enrollment presents a somewhat different story [Column (3)], children who do not speak Spanish are only 15.2 percent less likely to be enrolled relative to bilingual children if there is an indigenous school in their community. This gap contrasts with an 18.1 negative probability of monolingual children who only have access to Spanish schools. This difference-in-difference, while implying a 2.9 percent reduction in the language barrier gap, is however, not significant, given our large sample size.[22]

An obvious empirical problem which arises here is that of endogenous program placement (Rosenzweig and Wolpin, 1986). It is likely that the Mexican government locates indigenous schools precisely in areas with higher indigenous population and where the indigenous tend to be less integrated, more isolated and consequently less likely to learn Spanish. If this is the case, the difference-in-difference OLS estimators

[20] A key underlying assumption in this matching procedure is that primary-level children do not attend community schools other than their own, since this would prevent us from matching the *true* underlying community school infrastructure to the corresponding child. We believe, however, that *conditional on the community having a primary school*, children may not choose to commute to other communities, given that our sample is characterized by households with very low resources located in relatively isolated communities. Another potential problem is that due to migration, some children over their life course may have attended to different schools other than the community school where they live at the time of the interview. The cross-section design of our sample prevents us to correct for this problem, and consequently the effect of community educational infrastructure on long term schooling outcomes may be biased.

[21] The negative effect of speaking an indigenous language is larger in this set of regressions than in the results reported in Table 3 and 4, reflecting the somewhat selected nature of our sample. This sample has omitted children living in communities which have more than one primary school, as well as communities with no high schools. It also focuses only on children in primary school.

[22] Given the large sample sizes, it may be appropriate to adopt a Bayesian approach to model selection. Following Schwarz (1978), the *a posteriori* most likely model will be chosen if a t statistics greater than

would be biased downward. To correct for possible endogenous program placement, we control for observed and unobserved community heterogeneity and re-estimate model (3) using community fixed effects. As we have expected, Columns (2) and (4) show that once we control for community fixed effects, the difference-in-difference coefficients increase in magnitude for both of our schooling outcomes. In terms of school enrollment, the positive sign of the interaction of the child being monolingual with the presence of an indigenous school, suggests that bilingual schools are able to reduce the schooling disadvantage by reducing the language barrier between monolingual and bilingual indigenous children. It is not surprising that the reduction is only significant for our short term schooling outcome, since indigenous schools are a relatively new concept in education in Mexico. Additionally, for years of completed schooling, we have no way of controlling for whether a child has been a resident in the current community and therefore likely to attend to the school we are attributing since he/she began to attend school. Thus, our empirical test is much cleaner using the variable school enrollment.

These results suggest that indigenous primary schools may help reduce the differences in educational attainment between monolingual and bilingual children, but the reader might view this as of questionable use if indigenous primary schools had the overall effect of reducing the educational attainment of both groups. This, however, does not appear to be the case. Our OLS estimations in Column (1) of Table 5, suggest that the level effect, -- e.g. the overall effect of having an indigenous primary school -- is positive and significant, despite the coefficients' potential downward bias.

## 6. Conclusions.

While it is routinely believed that the indigenous population tends to be among the poorest in terms of income or consumption measures in Mexico, there has thus far been little evidence on the educational attainment of indigenous children. This paper has provided a first step towards a diagnostic of the factors affecting educational attainment of indigenous children.

---

3.17 is judged significant in the regressions in the table.

We have shown that indigenous children on average fair worse in educational outcomes than non-indigenous children, even within highly marginated rural areas of Mexico. Nevertheless, we also show that there is great heterogeneity within the indigenous population. When indigenous children also learn Spanish, they achieve educational outcomes which are almost equivalent to their non-indigenous counterparts. When they do not learn Spanish, nevertheless, their educational outcomes are far inferior.

Our analysis has shed some light on the explanatory factors for why some monolingual indigenous children do worse off than bilingual indigenous children. Instrumental variable procedures and evidence on schooling outcomes in bilingual educational programs, suggest that the language barrier for children who do not speak Spanish is an important factor that prevents them to achieve high schooling outcomes.

Bilingual education is a relatively new phenomenon in education in Mexico. Our results are suggestive of potential positive effects of indigenous primary schools. That is, indigenous children appear to enroll more in school when there is an indigenous primary located in their community and the negative gap of language is substantially reduced. Nevertheless the impacts and results of bilingual schools is an important area for evaluation which should continue to be monitored. It is clearly still too early to speculate on its long-term effects. Within the context of bilingual schooling in the United States, Duignan (2000) argues that there is strong evidence that bilingual schooling reduces the probability that children learn English and reduces assimilation rates. Therefore, one area of possible concern is the impact of bilingual education on indigenous children's learning of Spanish. Positive impacts of bilingual education could be undermined if bilingual education reduces the probability of indigenous children learning Spanish. Important also, is that bilingual education is limited to pre-school and primary education in Mexico. Bilingual programs do not currently exist at the secondary school level.

Finally, while we have emphasized the important effect of language in this paper, economic conditions also explain a large portion of educational differences between

indigenous and non-indigenous children. Thus, social programs to improve the economic conditions of indigenous households will also improve the educational attainment of indigenous children.

**References**


Aguirre Beltran, G. (1967). "Regiones de refugio: El desarrollo de la comunidad y el proceso dominical en mestizoamerica. Instituto Nacional Indigenista." *Fondo de Cultura Económica.*

Aguirre Beltran, G. (1970). "El Proceso de Aculturización y el cambio socio-cultural en México." *Instituto Nacional Indigenista. Fondo de Cultura Económica.*

Barro, R. and Lee, J. (1996) "International Measures of Schooling Years and Schooling Quality." *American Economic Review, Papers and Proceedings, 86:218-223.*

Basmann, R., (1960). "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association, 55:650-659.*

Behrman, J. and J. Knowles (1999). "Household Income and Child Schooling in Vietnam". *World Bank Economic Review.* Vol. 13:2. 211-156.

Behrman, J. and P. Todd. (1999). "Randomness in the Experimental Samples of Progresa (Education, Health and Nutrition Program)." *Progresa evaluation documents.*

Behrman, J., P. Sengupta and P. Todd. (2000). "Final Report: The Impact of Progresa on Achievement Scores in the First Year." *Progresa evaluation documents.*

Behrman, J., N. Birdsall, and M. Szekeley. (1999). "Intergenerational Mobility in Latin America: Deeper Markets and Better Schools Make a Difference" in Nancy Birdsall and Carol Graham, eds., New Markets, New Opportunities? Economic and Social Mobility in a Changing World, Washington, DC: The Brookings Institution and the Carnegie Endowment for International Peace.

Blake, J. (1989). "Family Size and Achievement." *Berkeley: University of California Press.*

Caso, A., S. Zavala, J. Miranda and M. González Navarro. (1981). "La Politica Indigenista en Mexico, Metodos y resultados."

Chiswick, B.R. and P.W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses" Journal of Labor Economics. Vol. 13:2.pp. 246-288.

Church, J. and I. King. 1993. "Bilingualism and network externalities" Canadian Journal of Economics. XXVI:2. p. 337-345.

Duignan, Peter. (2000). "Bilingual Education: A Critique." *Mimeo Hoover Institution.*

Heer, David M. (1985). "Effects of Sibling Number on Child Outcome." *in Ralph H Turner and James F. Short, Jr. (eds) Annual Review of Sociology, 11: 27-47. Palo Alto: Reviews Inc.*

Hausman, J.A., (1978). "Specification Tests in Econometrics." *Econometrica* 46:1251-1272.

Hauser, R. M. and Daphne Kuo Hsiang-Hui (1995). "Does the Gender Composition of Sibships affect Educational Attainment?" *CDE Working Paper 95-06*.

Hernsein, Richard J. and Charles Murray, (1994). "The Bell Curve." *New York: Free Press.*

INEGI (1994). "Estadísticas Históricas de México."

INEGI (2000). "Información preliminar, Censo Nacional de Población 2000."

John, A., and K. Yi. 1997. "Language, Learning and Location". Mimeo. University of Virginia.

Lang, K. 1986. "A Language Theory of Discrimination." Quarterly Journal of Economics. Pp. 363-382.

Lazear, E. P. 1999. "Culture and Language." *Journal of Political Economy*. 107: 95-126.

Lombardo Toledano, V. 1991. "Escritos Acerca de la Situacion de los Indigenas." *Centro de Estudios Filosoficos, Politicos y Sociales.*

López Acevedo, Gladys. (1999). "Learning Outcomes and School Cost-Effectiveness in Mexico". *Policy Research Working Paper 2128. The World Bank.*

Modiano, N (1988). "Public Bilingual Education in Mexico." *In C.B. Paulston, ed., Handbook of Bilingualism and Bilingual Education. New York, Greenwoood Press.*

Palafox, Juan Carlos, Juan Prawda, and Eduardo Velez. (1994). "Primary School Quality in Mexico". *Comparative Education Review. 38, no. 2.*

Panagides, Alexis. (1994). "Mexico." In *Indigenous People and Poverty in Latin America. An Empirical Analysis, by George Psacharopoulos and Harry A. Patrinos, World Bank.*

Pool, J. 1991. "The Official Language Problem." American Political Science Review. Vol. 85:2. pp. 495-514.

Robichaux. D. (1994). "Clase, percepción étnica y transformación regional: unos ejemplos tlaxcaltecas" *Boletín de Antropología Americana. 30.*

Schultz, T.P. (2000). "School Subsidies for the Poor: Evaluating a Mexican Strategy for Reducing Poverty." *Mimeo*.

Scott, J. (2000). "Who Benefits from the State in High-Inequality, Middle-Income Countries? The Case of Mexico." *CIDE mimeo.*

Secretaria de Eduacación Pública. (2000). "Perfil de la Eduacación en México."

Schwartz, G. (1978). "Estimating the dimension of a model. "*Annals of Statistics, 6:461-64.*

Figure 1

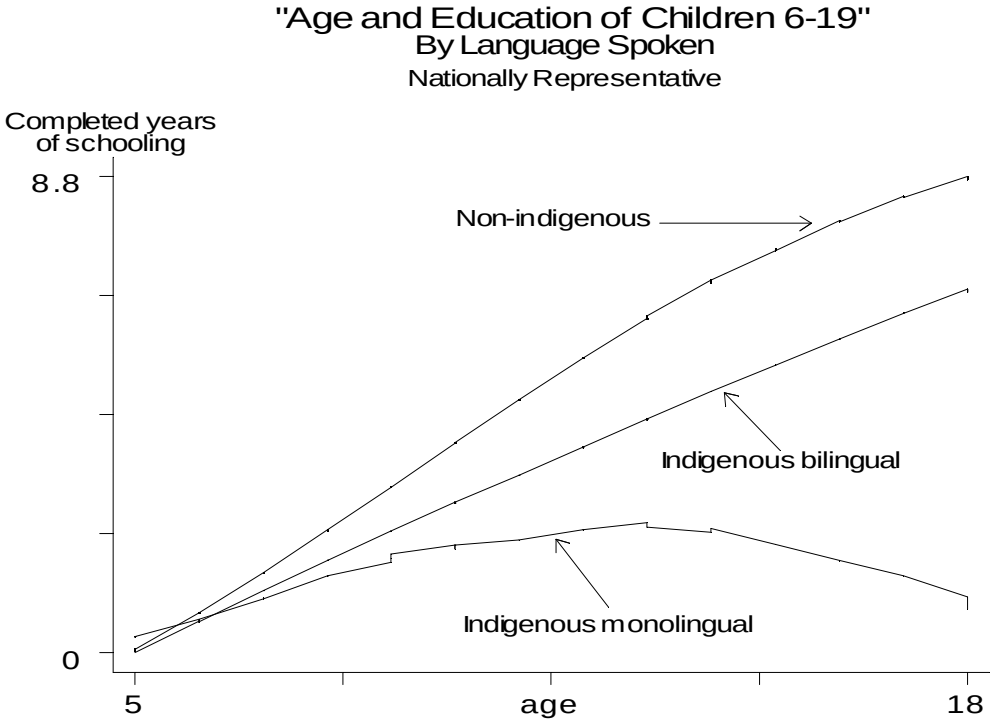



Figure 2

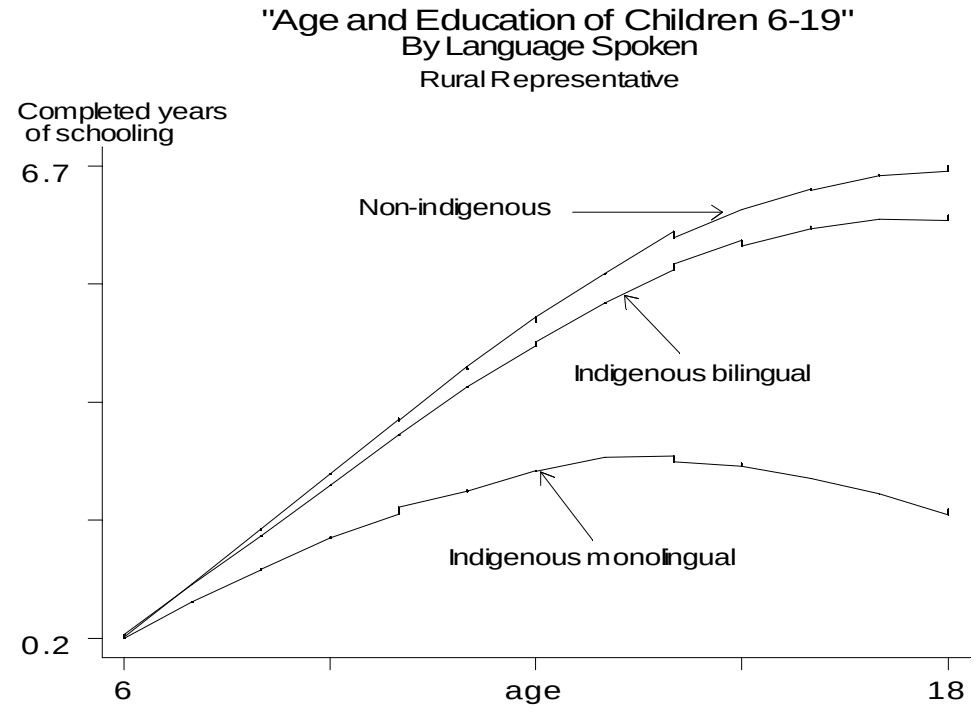


Figure 3

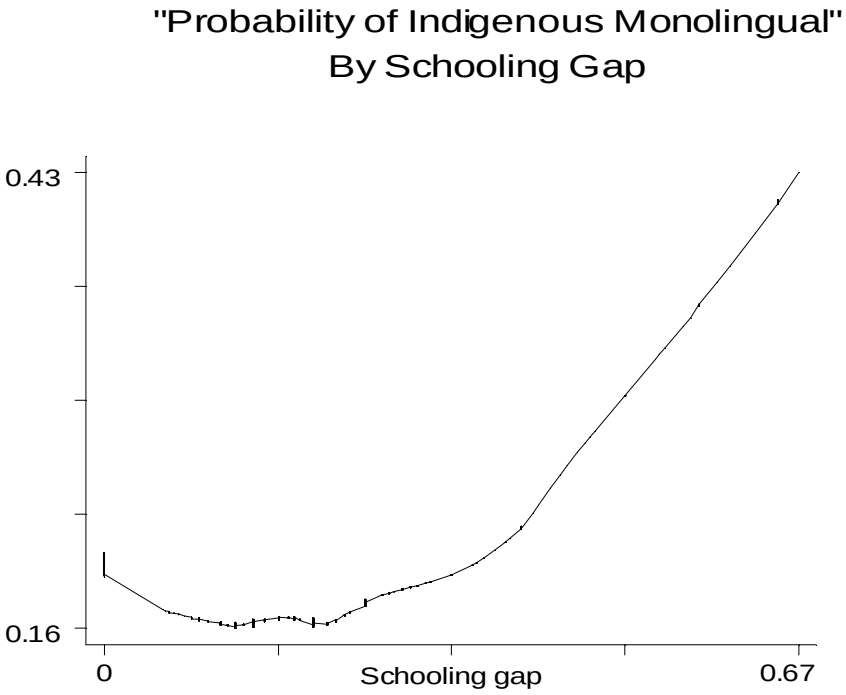

"Probability of Indigenous Monolingual"
By Schooling Gap
0.43
0.16
0
Schooling gap
0.67

Table 1
Determinants of years of completed schooling: the effect of being indigenous children aged 6 to 18

| | OLS | | | | | CFE | |
|---|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
| **Child Characteristics** | | | | | | | |
| Child is indigenous | -0.571 | -0.58 | -0.372 | -0.348 | -0.246 | -0.253 | -0.501 |
| | [0.010]*** | [0.010]*** | [0.010]*** | [0.010]*** | [0.011]*** | [0.025]*** | [0.234]** |
| Gender (Boy=1) | 0.002 | -0.007 | -0.003 | -0.003 | -0.004 | -0.002 | 0.043 |
| | [0.008] | [0.008] | [0.008] | [0.008] | [0.008] | [0.008] | [0.054] |
| Age 9 to 11 | 2.203 | 2.205 | 2.259 | 2.214 | 2.203 | 2.18 | 2.321 |
| | [0.008]*** | [0.008]*** | [0.008]*** | [0.008]*** | [0.008]*** | [0.011]*** | [0.078]*** |
| Age 12 to 14 | 4.227 | 4.233 | 4.341 | 4.254 | 4.232 | 4.191 | 4.385 |
| | [0.010]*** | [0.010]*** | [0.010]*** | [0.010]*** | [0.010]*** | [0.012]*** | [0.081]*** |
| Age 15 to 18 | 5.324 | 5.354 | 5.505 | 5.374 | 5.347 | 5.326 | 5.577 |
| | [0.012]*** | [0.013]*** | [0.012]*** | [0.012]*** | [0.012]*** | [0.012]*** | [0.083]*** |
| **Parental Characteristics** | | | | | | | |
| Father's age | . | -0.002 | . | 0.008 | 0.007 | 0.003 | 0 |
| | | [0.001]** | | [0.001]*** | [0.001]*** | [0.001]*** | [0.005] |
| Mother's age | . | 0.002 | . | 0.011 | 0.009 | 0.008 | 0.013 |
| | | [0.001]*** | | [0.001]*** | [0.001]*** | [0.001]*** | [0.006]** |
| Father's edu 1 to 5 years | . | . | 0.398 | 0.434 | 0.406 | 0.318 | 0.266 |
| | | | [0.012]*** | [0.012]*** | [0.012]*** | [0.014]*** | [0.095]*** |
| Father's edu 6 $^{+}$ years | . | . | 0.656 | 0.748 | 0.645 | 0.425 | 0.288 |
| | | | [0.014]*** | [0.014]*** | [0.014]*** | [0.017]*** | [0.111]*** |
| Mother's edu 1 to 5 years | . | . | 0.484 | 0.522 | 0.479 | 0.37 | 0.333 |
| | | | [0.011]*** | [0.011]*** | [0.011]*** | [0.013]*** | [0.091]*** |
| Mother's edu 6 $^{+}$ years | . | . | 0.792 | 0.879 | 0.746 | 0.5 | 0.722 |
| | | | [0.013]*** | [0.013]*** | [0.013]*** | [0.017]*** | [0.107]*** |
| **Assets** | | | | | | | |
| Cement floor | . | . | . | . | 0.306 | 0.258 | 0.292 |
| | | | | | [0.012]*** | [0.016]*** | [0.088]*** |
| Hhold has water and electricity | . | . | . | . | 0.302 | 0.233 | 0.158 |
| | | | | | [0.013]*** | [0.016]*** | [0.083]* |
| Hhold owns agric. land | . | . | . | . | 0.034 | 0.154 | 0.257 |
| | | | | | [0.009]*** | [0.011]*** | [0.081]*** |
| Hhold has refrig. and stove | . | . | . | . | 0.469 | 0.486 | 0.367 |
| | | | | | [0.011]*** | [0.015]*** | [0.075]*** |
| Observations | 220,008 | 220,008 | 220,008 | 220,008 | 220,008 | 220,008 | 4,640 |
| R-squared | 0.53 | 0.53 | 0.55 | 0.55 | 0.56 | 0.57 | 0.59 |
| Number of communities | . | . | . | . | . | 25,905 | 255 |

Notes: Columns (1) through (6) present results using the Encaseh 32-state sample. Column (7) estimates correspond to the rural national sample. Robust standard errors in [brackets]. Coefficients marked with (***) are significant under Schwartz(1978) a posteriori criteria, where the most likely model is chosen with a t statistic no smaller than 3.5.

Table 2
Long versus short term schooling outcomes

| | Years of schooling | | School enrollment | | | |
|---|---|---|---|---|---|---|
| | OLS | CFE | OLS | CFE | OLS | CFE |
| | [5] | [6] | [1] | [2] | [3] | [4] |
| **Child Characteristics** | | | | | | |
| Child is indigenous | -0.246 | -0.253 | 1.81 | -3.227 | . | . |
| | [0.011]*** | [0.025]*** | [0.205]*** | [0.501]*** | | |
| Only speaks indigenous language | . | . | . | . | -10.662 | -17.036 |
| | | | | | [0.454]*** | [0.725]*** |
| Speaks indigenous & Spanish | . | . | . | . | 4.334 | -2.664 |
| | | | | | [0.214]*** | [0.501]*** |
| Gender (Boy=1) | -0.004 | -0.002 | 3.353 | 3.431 | 3.253 | 3.344 |
| | [0.008] | [0.008] | [0.159]*** | [0.162]*** | [0.159]*** | [0.162]*** |
| Age 9 to 11 | 2.203 | 2.18 | 5.319 | 4.985 | 4.884 | 4.601 |
| | [0.008]*** | [0.011]*** | [0.179]*** | [0.230]*** | [0.178]*** | [0.230]*** |
| Age 12 to 14 | 4.232 | 4.191 | -14.386 | -14.635 | -15.059 | -15.225 |
| | [0.010]*** | [0.012]*** | [0.241]*** | [0.236]*** | [0.240]*** | [0.237]*** |
| Age 15 to 18 | 5.347 | 5.326 | -57.617 | -57.304 | -58.378 | -58.001 |
| | [0.012]*** | [0.012]*** | [0.245]*** | [0.240]*** | [0.243]*** | [0.241]*** |
| **Parental Characteristics** | | | | | | |
| Father's age | 0.007 | 0.003 | 0.056 | -0.002 | 0.048 | -0.004 |
| | [0.001]*** | [0.001]*** | [0.013]*** | [0.015] | [0.013]*** | [0.015] |
| Mother's age | 0.009 | 0.008 | 0.003 | -0.014 | -0.004 | -0.013 |
| | [0.001]*** | [0.001]*** | [0.014] | [0.016] | [0.014] | [0.016] |
| Father's edu 1 to 5 years | 0.406 | 0.318 | 4.968 | 4.086 | 4.642 | 4.07 |
| | [0.012]*** | [0.014]*** | [0.245]*** | [0.283]*** | [0.244]*** | [0.283]*** |
| Father's edu 6 $^{+}$ years | 0.645 | 0.425 | 9.45 | 7.196 | 8.941 | 6.94 |
| | [0.014]*** | [0.017]*** | [0.289]*** | [0.350]*** | [0.289]*** | [0.349]*** |
| Mother's edu 1 to 5 years | 0.479 | 0.37 | 6.073 | 4.989 | 5.542 | 4.898 |
| | [0.011]*** | [0.013]*** | [0.221]*** | [0.259]*** | [0.221]*** | [0.259]*** |
| Mother's edu 6 $^{+}$ years | 0.746 | 0.5 | 10.957 | 7.669 | 10.369 | 7.515 |
| | [0.013]*** | [0.017]*** | [0.269]*** | [0.332]*** | [0.269]*** | [0.332]*** |
| **Assets** | | | | | | |
| Cement floor | 0.306 | 0.258 | 2.473 | 3.317 | 2.401 | 3.308 |
| | [0.012]*** | [0.016]*** | [0.241]*** | [0.321]*** | [0.241]*** | [0.321]*** |
| Hhold has water and electricity | 0.302 | 0.233 | 3.135 | 2.768 | 3.095 | 2.766 |
| | [0.013]*** | [0.016]*** | [0.243]*** | [0.314]*** | [0.242]*** | [0.314]*** |
| Hhold owns agric. land | 0.034 | 0.154 | 1.473 | 2.463 | 1.694 | 2.47 |
| | [0.009]*** | [0.011]*** | [0.167]*** | [0.229]*** | [0.167]*** | [0.228]*** |
| Hhold has refrig. and stove | 0.469 | 0.486 | 5.471 | 5.329 | 5.579 | 5.372 |
| | [0.011]*** | [0.015]*** | [0.215]*** | [0.292]*** | [0.215]*** | [0.292]*** |
| Observations | 220,008 | 220,008 | 220,712 | 220,712 | 220,716 | 220,716 |
| R-squared | 0.56 | 0.57 | 0.34 | 0.35 | 0.35 | 0.35 |
| Number of communities | . | 25,905 | . | 25,907 | . | 25,907 |

Notes: See table one. Results using the Encaseh 32 state-sample. School enrollment indicator multiplied by 100.

Table 3
Long versus short term schooling outcomes for monolingual and bilingual indigenous children

| | Years of schooling | | | | | School enrollment | |
|---|---|---|---|---|---|---|---|
| | **OLS** | **CFE** | **CFE** | **CFE** | **IV** | **CFE** | **IV** |
| | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
| **Child Characteristics** | | | | | | | |
| Only speaks indigenous language | -1.455 | -1.073 | . | -1.091 | -0.942 | -14.255 | -9.926 |
| | [0.023]*** | [0.029]*** | | [0.030]*** | [0.055]*** | [0.580]*** | [1.146]*** |
| Gender (Boy=1) | 0.193 | 0.189 | 0.214 | 0.188 | 0.181 | 5.992 | 5.511 |
| | [0.018]*** | [0.017]*** | [0.017]*** | [0.017]*** | [0.017]*** | [0.342]*** | [0.339]*** |
| Age 9 to 11 | 1.786 | 1.799 | 1.912 | 1.797 | 1.848 | 4.008 | 5.081 |
| | [0.016]*** | [0.024]*** | [0.024]*** | [0.024]*** | [0.017]*** | [0.485]*** | [0.425]*** |
| Age 12 to 14 | 3.479 | 3.518 | 3.694 | 3.515 | 3.593 | -14.758 | -13.685 |
| | [0.021]*** | [0.025]*** | [0.025]*** | [0.025]*** | [0.023]*** | [0.502]*** | [0.551]*** |
| Age 15 to 18 | 4.302 | 4.414 | 4.618 | 4.411 | 4.49 | -56.786 | -55.741 |
| | [0.026]*** | [0.026]*** | [0.025]*** | [0.026]*** | [0.028]*** | [0.514]*** | [0.566]*** |
| **Parental Characteristics** | | | | | | | |
| Father's age | . | 0.003 | 0.004 | 0.003 | 0.007 | 0.036 | 0.071 |
| | | [0.002]* | [0.002]** | [0.002]* | [0.001]*** | [0.031] | [0.027]*** |
| Mother's age | . | 0.005 | 0.006 | 0.005 | 0.008 | -0.073 | -0.002 |
| | | [0.002]*** | [0.002]*** | [0.002]*** | [0.002]*** | [0.033]** | [0.030] |
| Father's edu 1 to 5 years | . | 0.321 | 0.331 | 0.33 | 0.414 | 4.438 | 5.772 |
| | | [0.027]*** | [0.029]*** | [0.028]*** | [0.023]*** | [0.543]*** | [0.478]*** |
| Father's edu 6 $^{+}$ years | . | 0.409 | 0.437 | 0.42 | 0.628 | 8.317 | 10.626 |
| | | [0.036]*** | [0.038]*** | [0.037]*** | [0.029]*** | [0.726]*** | [0.597]*** |
| Mother's edu 1 to 5 years | . | 0.213 | 0.22 | 0.23 | 0.415 | 3.96 | 6.295 |
| | | [0.026]*** | [0.028]*** | [0.027]*** | [0.021]*** | [0.524]*** | [0.438]*** |
| Mother's edu 6 $^{+}$ years | . | 0.283 | 0.32 | 0.304 | 0.604 | 5.281 | 9.258 |
| | | [0.038]*** | [0.040]*** | [0.039]*** | [0.029]*** | [0.768]*** | [0.602]*** |
| Father only speaks indigenous language | . | . | -0.13 | 0.039 | . | . | . |
| | | | [0.041]*** | [0.041] | | | |
| Mother only speaks indigenous language | . | . | -0.115 | 0.069 | . | . | . |
| | | | [0.032]*** | [0.032]** | | | |
| **Assets** | | | | | | | |
| Cement floor | . | 0.306 | 0.323 | 0.305 | 0.371 | 2.827 | 1.852 |
| | | [0.047]*** | [0.048]*** | [0.047]*** | [0.040]*** | [0.950]*** | [0.745]** |
| Hhold has water and electricity | . | 0.194 | 0.202 | 0.194 | 0.119 | 2.843 | 1.033 |
| | | [0.039]*** | [0.040]*** | [0.039]*** | [0.032]*** | [0.787]*** | [0.631] |
| Hhold owns agric. land | . | 0.158 | 0.162 | 0.159 | 0.14 | 3.022 | 4.189 |
| | | [0.026]*** | [0.026]*** | [0.026]*** | [0.020]*** | [0.518]*** | [0.397]*** |
| Hhold has refrig. and stove | . | 0.579 | 0.602 | 0.583 | 0.748 | 5.905 | 7.361 |
| | | [0.059]*** | [0.060]*** | [0.059]*** | [0.056]*** | [1.182]*** | [0.939]*** |
| **Hypothesis Testing** | | | | | | | |
| IV overidentification test $X^2_{[21]}$ | . | . | . | . | 0.14 | . | 0.00 |
| | | | | | (1.000) | | (1.000) |
| $R^2$ First-stage regression | . | . | . | . | 0.27 | . | 0.27 |
| Hausman test for exogeneity $X^2_{[21]}$ | . | . | . | . | 42.04 | . | 23.97 |
| | | | | | (0.004) | | (0.244) |
| Observations | 51,229 | 51,229 | 51,229 | 51,229 | 51,229 | 51,492 | 51,492 |
| R-squared | 0.47 | 0.5 | 0.49 | 0.5 | 0.5 | 0.34 | 0.33 |
| Number of communities | . | 6,432 | 6,432 | 6,432 | . | 6,433 | . |

Notes: Results using Encaseh 32-states sample *for indigenous children only* . School enrollment indicator multiplied by 100. Robust standard errors in [brackets]. Test p-values in (parenthesis). Coefficients marked with (***) are significant under Schwartz' (1978) , where the most likely model is chosen with a t statistic no smaller than 3.3.

Table 4

Estimated effects of being monolingual on schooling indicators of indigenous children: by age
Instrumental variable estimations

| | Age group | | | |
|---|---|---|---|---|
| | **6 to 8** | **9 to 11** | **12 to 14** | **15 to 18** |
| **Years of schooling** | | | | |
| Child only speaks indigenous language | -0.069<br>[0.030]** | -0.414<br>[0.071]*** | -1.602<br>[0.132]*** | -2.265<br>[0.213]*** |
| Observations | 12,638 | 12,227 | 12,236 | 14,128 |
| R-squared | 0.33 | 0.22 | 0.21 | 0.24 |
| **School enrollment** | | | | |
| Child only speaks indigenous language | -10.586<br>[1.555]*** | -9.364<br>[1.712]*** | -15.181<br>[2.967]*** | -4.268<br>[3.434] |
| Observations | 12,967 | 12,222 | 12,226 | 14,076 |
| R-squared | 0.09 | 0.07 | 0.12 | 0.13 |

Notes: Results using Encaseh 32-states sample for indigenous children only. School enrollment indicator multiplied by 100. Robust standard errors in [brackets]. Test p-values in (parenthesis). All models include child characteristics: sex and age; parental characteristics: father's age, mother's age, father and mother education groups; household assets: cement floor, water and electricity, agricultural land owning, refrigerator and stove. Fulll estimation results available from authors upon request.

Table 5

Long versus short term schooling outcomes for monolingual and bilingual indigenous children. Language and primary school-type interactions

| | Years of schooling | | School enrollment | |
|---|---|---|---|---|
| | **OLS** | **CFE** | **OLS** | **CFE** |
| | [1] | [2] | [3] | [4] |
| Child only speaks indigenous language | -0.846 | -0.975 | -18.128 | -18.384 |
| | [0.039]*** | [0.048]*** | [1.153]*** | [1.223]*** |
| Indigenous primary school | 0.089 | . | 1.639 | . |
| | [0.020]*** | . | [0.468]*** | . |
| Only speaks indigenous language interacted with indigenous school | 0.019 | 0.118 | 2.898 | 3.915 |
| | [0.046] | [0.059]** | [1.356]** | [1.485]*** |
| Observations | 23,972 | 23,972 | 23,833 | 23,833 |
| R-squared | 0.39 | 0.42 | 0.39 | 0.35 |
| Number of communities | . | 3,973 | . | 3,968 |

Notes: Results using Encaseh 32-states sample for indigenous children *between 6 to 12 years old in communities with only one primary school*. Robust standard errors in [brackets]. Coefficients marked with (***) are significant under Schwartz' (1978) with a t statistic no smaller than 3.17. All models include child characteristics: sex and age groups; parental characteristics: father's age, mother's age, father and mother education groups; household assets: cement floor, water and electricity, agricultural land owning, refrigerator and stove.

Appendix

## Table 1 A
## Descriptive Statistics "National and 32 State Samples"

| Variable | Mean National sample [Std. Err. ] | Mean 32 State sample [Std. Err. ] | Difference [Std. Err. ] |
|---|---|---|---|
| Age | 11.988 | 11.726 | 0.262 |
| | [0.032] | [0.008] | [0.033] |
| Gender (Boy=1) | 0.505 | 0.506 | -0.001 |
| | [0.004] | [0.001] | [0.004] |
| School enrollment | 78.552 | 69.151 | 9.402 |
| | [0.353] | [0.098] | [0.366] |
| Years of schooling | 5.021 | 3.941 | 1.08 |
| | [0.028] | [0.006] | [0.029] |
| Household size | 6.178 | 7.013 | -0.835 |
| | [0.019] | [0.005] | [0.020] |
| Only speaks indigenous language | 0.014 | 0.043 | -0.029 |
| | [0.001] | [0.000] | [0.001] |
| Speaks indigenous & Spanish | 0.033 | 0.191 | -0.158 |
| | [0.002] | [0.001] | [0.002] |
| Only speaks Spanish | 0.953 | 0.766 | 0.187 |
| | [0.002] | [0.001] | [0.002] |
| Father's education | 6.285 | 3.241 | 3.045 |
| | [0.035] | [0.006] | [0.036] |
| Mother's education | 5.718 | 2.844 | 2.874 |
| | [0.034] | [0.006] | [0.035] |
| Father's age | 41.709 | 42.886 | -1.178 |
| | [0.071] | [0.020] | [0.074] |
| Mother's age | 38.074 | 38.517 | -0.442 |
| | [0.067] | [0.018] | [0.069] |
| Cement floor | 0.327 | 0.121 | 0.206 |
| | [0.004] | [0.001] | [0.004] |
| Hhold has water and electricity | 0.563 | 0.126 | 0.437 |
| | [0.004] | [0.001] | [0.004] |
| Hhold owns agric. land | 0.183 | 0.564 | -0.38 |
| | [0.003] | [0.001] | [0.003] |
| Hhold has refrig. and stove | 0.615 | 0.181 | 0.434 |
| | [0.004] | [0.001] | [0.004] |
| Observations | 13,697 | 222,601 | |

Table 2A
Despcriptive Statistics National Sample (By Sector)

| Variable | Mean Urban [Std. Err. ] | Mean Rural [Std. Err. ] | Difference [Std. Err. ] |
|---|---|---|---|
| Age | 12.123 | 11.73 | -0.393 |
| | [0.040] | [0.054] | [0.067] |
| Gender (Boy=1) | 0.507 | 0.502 | -0.005 |
| | [0.005] | [0.007] | [0.009] |
| School enrollment | 83.121 | 69.852 | -13.269 |
| (percentage) | [0.398] | [0.672] | [0.781] |
| Years of schooling | 5.492 | 4.123 | -1.369 |
| | [0.036] | [0.043] | [0.056] |
| Household size | 5.778 | 6.939 | 1.161 |
| | [0.021] | [0.036] | [0.041] |
| Only speaks indigenous language | 0 | 0.039 | 0.038 |
| | [0.000] | [0.003] | [0.003] |
| Speaks indigenous & Spanish | 0.01 | 0.078 | 0.068 |
| | [0.001] | [0.004] | [0.004] |
| Only speaks Spanish | 0.99 | 0.883 | -0.107 |
| | [0.001] | [0.005] | [0.005] |
| Father's education | 7.347 | 4.266 | -3.081 |
| | [0.044] | [0.047] | [0.064] |
| Mother's education | 6.73 | 3.793 | -2.937 |
| | [0.042] | [0.046] | [0.062] |
| Father's age | 41.587 | 41.94 | 0.353 |
| | [0.086] | [0.127] | [0.153] |
| Mother's age | 38.252 | 37.736 | -0.516 |
| | [0.082] | [0.113] | [0.140] |
| Cement floor | 0.389 | 0.208 | -0.181 |
| | [0.005] | [0.006] | [0.008] |
| Hhold has water and electricity | 0.727 | 0.25 | -0.477 |
| | [0.005] | [0.006] | [0.008] |
| Hhold owns agric. land | 0.055 | 0.427 | 0.371 |
| | [0.002] | [0.007] | [0.008] |
| Hhold has refrig. and stove | 0.752 | 0.354 | -0.398 |
| | [0.005] | [0.007] | [0.008] |
| Observations | 13,697 | | |