Inter-American Development Bank
Banco Interamericano de Desarrollo
Latin American Research Network
Red de Centros de Investigación
Research Network Working paper #R-438

Schooling Inequality among the Indigenous: A Problem of Resources or Language Barriers?

by

Susan W. Parker*
Luis Rubalcava*
Graciela Teruel**

*Centro de Investigación y Docencia Económicas
**Universidad Iberoamericana

January 2002

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Parker, Susan Wendy.
Schooling inequality among the indigenous : a problem of resources or language barriers / by Susan W. Parker, Luis Rubalcava, Graciela Teruel.

p. cm. (Research Network working papers ; R-438)
Includes bibliographical references.

1. Educational equalization--Mexico. 2. Education, Bilingual--Mexico. 3. Indigenous peoples--Mexico. I. Rubalcava, Luis N. II. Teruel, Graciela M., 1968-. III. Inter-American Development Bank. Research Dept. IV. Latin American Research Network. V. Title. VI. Series.

370.19342 P331--dc21

The Research Department (RES) produces the *Latin American Economic Policies Newsletter*, as well as working papers and books, on diverse economic issues. To obtain a complete list of RES publications, and read or download them please visit our web site at: http://www.iadb.org/res

Abstract[*]

This paper analyzes an important determinant of social exclusion of the indigenous in Mexico, that of educational attainment. Using large household data sets from rural Mexican communities, where a majority of indigenous people live, 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) perform much worse in school than bilingual indigenous children who speak Spanish as a second language.

Using community effects and instrumental variable models that 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 characteristics, household wealth 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 and monolingual indigenous children. Rather, language barriers and cultural factors represent an important aspect of barriers that indigenous children face in school and thus an important determinant of their future social exclusion.

In order to better understand the extent to which these language barriers affect indigenous children's schooling outcomes and how they may differ from other cultural unobserved factors, we exploit the fact that the Mexican public educational system carries out two parallel programs. Indigenous children, depending on their place of residence, have access to either traditional or bilingual primary schools. We thus examine the possible role of bilingual education in improving the educational performance of indigenous children. Using a difference-in-difference estimation that sweeps out most of the cultural unobserved heterogeneity across indigenous groups, our results demonstrate that bilingual primary schools narrow the gap in the educational performance of monolingual children relative to bilingual children.

We conclude that while higher poverty levels in Mexico contribute to differences between indigenous and non-indigenous educational outcomes, they do not explain all of the observed differences. The analysis of this paper indicates that a large fraction of this worse performance is due to a lack of knowledge of Spanish, suggesting that social programs aiming only at improving the marginality conditions of indigenous communities will be insufficient for eliminating schooling inequality among the indigenous population.

[*] The authors thank Jere Behrman, Alejandro Gaviria, Miguel Szekely and all participants of the IADB seminar "Social Exclusion: Who is in, Who is out and Why does it Matter?" for useful comments. We are also grateful to 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 Inter-American Development Bank.

Introduction

Over the years Mexico has experienced important advances in its social indicators. At the beginning of the twentieth century 80 percent of the population was illiterate and the average Mexican completed only three years of schooling.[1] Currently, however, the median of the population has reached the first year of secondary school and only one out of ten Mexicans is unable to read and write.[2]

These general increases in human capital formation have, nevertheless, been associated with little reduction in income inequality.[3] Social progress has been far from homogeneous. Whereas urban areas have to a large extent seen improvements in their social indicators, remote rural communities have continued to lag far behind urban areas. These isolated communities are the home of the vast majority of indigenous groups who remain, in many dimensions, socially excluded from the rest of the Mexican population. This social exclusion may manifest itself in many forms, from higher poverty rates to other forms of political and cultural exclusion.

This paper concentrates on one type of social exclusion that the indigenous face, that of access to the formal education system. Unlike the average Mexican, the indigenous have seen little improvement in their schooling performance over time. Educational attainment, furthermore, is likely the variable that will most determine the income and poverty levels of children in the future and is thus critical for avoiding the transmission of social exclusion to future indigenous generations. Evidence on the importance of the education of the indigenous is shown in Table 1, which compares the labor income of indigenous and non-indigenous prime age males. Overall, average labor earnings for the indigenous are only about half those of non-indigenous men. These differences, however, are greatest for individuals with low levels of schooling (primary or less), whereas adults with secondary education or higher show no significant differences in labor earnings between indigenous and non-indigenous men. The table implies that with increasing education levels the negative effect of being indigenous on earnings is greatly reduced, or even disappears. In this way, improving education levels of the indigenous can significantly reduce the degree of their social exclusion, at least in economic terms.

The potential problems and limitations which indigenous children may face in school is a subject on which little evidence exists either in the literature as a whole or in the specific case of

[1] INEGI (1994).
[2] Barro and Lee (1996) and Scott (2000).

Mexico. This situation exists in spite of the fact that the size of the indigenous population approaches 10 million—the largest indigenous population in Latin America. The current paper analyzes to what extent being indigenous is associated with lower schooling outcomes for children living in remote rural communities. The paper compares different schooling indicators among children with the same socioeconomic opportunities, household demographics and community characteristics, age and gender, but with different ethnic backgrounds. The analysis exploits unique survey information on households in Mexico's poor rural communities, where most of the country's indigenous households are located. The paper also uses a nationally representative household survey to carry out the same analysis to investigate the comparability of the results.

This paper represents 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 low performance have not been studied. Potential explanatory factors include lower family resources, access to lower quality schools, discrimination as well as cultural and language barriers (Caso, Zavala, Miranda *et al*., 1981). If indigenous households are poorer than other households and poverty is a factor affecting school decisions, then simple correlations cannot determine whether 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. This paper attempts to shed some light on the extent to which family resources versus cultural factors and versus language barriers are related to the lower educational outcomes of indigenous children. This also makes it possible to analyze the extent to which indigeneity is a heterogeneous concept, that is, whether some indigenous children are more socially excluded than others and why.

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

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

Regression models of the determinants of children's schooling outcomes provide concrete evidence as to the reasons for this poor performance. The models first control only for background family characteristics and resources at the household level, followed by community effects. Instrumental variable methods are then used to explore the possible endogeneity of language spoken within the indigenous population. While controlling for parental characteristics, household wealth and community variables reduces the overall size of the negative effect of speaking only a native language, the effect remains significant. These results indicate that while family and community resources are clearly important, they cannot explain all of the differences in educational attainment between bilingual indigenous and monolingual indigenous children. Rather, language barriers or cultural factors represent an important aspect of problems that indigenous children face in school.[4] Indigenous children who do not learn Spanish are likely to face greater social exclusion both in and out of school than their bilingual counterparts.

In order to better understand the extent to which these language barriers affect schooling outcomes and how they differ from other cultural unobserved factors, the study exploits the fact that the Mexican public education system offers bilingual education at the primary level in a system parallel system to its traditional schools. The Secretary of Public Education (SEP) offers bilingual primary schools that include bilingual teachers as well as textbooks in native languages. It is thus possible to examine whether the availability of bilingual schools can reduce the language barriers described above. Based on a difference-in-difference estimation that sweeps out most of the cultural unobserved heterogeneity across indigenous groups, the results suggest that the educational disadvantage due to language barriers is reduced for children who have the option of attending a primary bilingual school. In this sense, having access to a bilingual school may reduce the social exclusion of monolingual indigenous children.

These results have important policy implications for indigenous learning in Mexico. Controlling for an important number of measures of family resources, access to schools and

[4] There are many potential factors that may affect investment (or lack thereof) of indigenous families in the education of their children. These factors may or may not relate to market failures. Possible market failures affecting this investment include the fact that indigenous families may have less information than non-indigenous families about the returns to education or less access to credit with which to finance their children's schooling. Alternatively, if the available quality of education to indigenous children is lower than that available to non-indigenous children, potential returns to indigenous education may be lower, and in this sense it may be "rational" or "efficient" for the indigenous to invest less. Or, indigenous parents may expect their children to work in activities with low returns to schooling and so may see the overall benefits to schooling as lower than non-indigenous parents. This paper does not attempt to distinguish between the efficiency and distributional aspects of lower educational attainment by indigenous children.

community characteristics, indigenous monolingual children continue to perform worse than their bilingual counterparts as well as non-indigenous children. This suggests that not knowing Spanish represents an important factor in their lower levels of performance. Indigenous primary bilingual schools, which practice bilingual education, improve the educational performance of monolingual children relative to bilingual children, evidence that corroborates the importance of language barriers

Thus, higher poverty levels can only partially explain differences between indigenous and non-indigenous educational outcomes, suggesting that social programs aiming only at lessening the marginality of indigenous communities will not be sufficient to eliminate schooling inequality among the indigenous. 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 bilingual primary schools seem to improve the performance of indigenous children, expansion of these integrated educational programs may be warranted. One possible caveat is that their effectiveness may be reduced if bilingual primary schools decrease the probability that indigenous children learn Spanish and the corresponding likelihood of assimilation. Stated another way, whereas bilingual schools may reduce the social exclusion of school-age indigenous children, the issue is whether they may eventually increase social exclusion of adults if bilingual schools reduce the probability of learning Spanish during childhood.

2. Background

There appear to be no previous studies that have analyzed the effect of indigenous background and language on children's schooling attainment. Perhaps the closest in nature is a study by Rosenthal, Baker and Ginsburg (1983) that examines the impact of language (English literacy) on children's achievement in math and reading in primary school in the United States. They find that household economic factors explain a majority of the estimated differences between the achievement of native English and Spanish- speaking children. Nevertheless, a sizable language barrier remains even after controlling for household economic variables.

There is also a literature in the United States on the impact of language spoken on the earnings of immigrants, which has some distant relation to the present study. For instance, Chiswick (1991) analyzes the determinants of English language fluency among migrants to the

United States and the relationship this fluency has with earnings. He shows that length of time in the United States is closely associated with English language fluency and that language fluency is an important determinant of earnings. In a subsequent paper, Chiswick and Miller (1995) attempt to treat language fluency as endogenous in the context of an analysis of the foreign born population in Australia, analyzing whether individuals who anticipate greater earnings from fluency are more likely to become fluent. They conclude that language fluency is in fact endogenous to earnings, although the exogeneity of their identifying instruments is somewhat questionable.

Even at a purely descriptive level, there has been little previous research on the educational attainment of indigenous children (in Mexico or elsewhere). One exception is Panagides (1994), who uses the Survey of Income and Expenditures (ENIGH) to look at various economic dimensions of indigenous individuals and families in Mexico, 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.

López (1999) has analyzed the impact of a Mexican social program called PARE (*Programa para abatir el Rezago Educativo*), which provided 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 bilingual schools. As a result of the program, improvements in test scores occurred in all areas, with the greatest improvements occurring in bilingual schools, although even after the program, test scores in bilingual 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 used in the present study.

Indigeneity is a complex subject, involving cultural traditions, languages and practices that 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 the definition of indigeneity would ideally be multi-dimensional, including not just language but other indicators as well, data constraints restrict the present definition to language spoken.

Fortunately, however, the available data include individual definitions of language spoken. Each individual is asked if he/she speaks an indigenous language. Those that report they do speak an indigenous language are then asked if they also speak Spanish. In this way the important distinction can be made between indigenous children who speak only a native language and indigenous children who are bilingual.[5]

It is also fortunate that there are 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, particularly between the father and mother. It is much more common for fathers to be bilingual than mothers.

This paper exploits the richness of individual level data to analyze the impact of being indigenous on children's schooling outcomes. Taking into account the variation that exists between mother, father and child languages spoken, it is possible to control for endogeneity of languages spoken within the indigenous population. That is, it is possible to take into account the fact that learning Spanish may be a choice.

3. Descriptive Analysis

The analysis begins 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 is located in rural areas. Within urban areas, only 1.1 percent of children are indigenous, versus 11.8 percent of children in rural areas.[6]

Using non-parametric regression[7] with a national sample of the Mexican population, Figure 1 illustrates the relationship between completed years of schooling for three groups of

[5] 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 due to possible ethnic discrimination concerns, a downward bias may result in the estimation of the differential schooling outcomes between non-indigenous and indigenous children whose school performance is the lowest.

[6] It should be noted that the national sample of the ENCASEH only contains information on 9,910 households, so the number of indigenous cases is quite small. The number of indigenous children between the ages of 6 and 18 is 658, and fewer than 200 report speaking only an indigenous language. Thus, the nationwide percentages of the indigenous population may differ compared with other nationally representative surveys, such as the Census. The nationally representative ENCASEH is used only to insure that the results are not overly biased by the use of a sample which is not representative at the national level.

[7]The non-parametric estimator applied carries out locally weighted, smoothed scatter plots (LOWESS). In this

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 bilingual children. Indigenous monolingual children however, lag behind in 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 entire child-age distribution.

Figure 2 shows the same graph as in Figure 1 but for a sample of marginalized rural areas, 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 twice that level, or on average about 7 years of schooling. The results from this regression foreshadow the regression analysis, which will demonstrate the great importance of language in determining educational outcomes.

4. Data Description

The data used for this project come directly from 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 by children and health clinic visits by the family. The program has collected a large quantity of socio-economic information as a result of both its mechanism of selection of beneficiaries and its evaluation.

---

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.

Two two principal, related sources of information are used for the analysis. The first is drawn from Progresa's targeting mechanism, which involves carrying out a socio-economic census (Survey of Household Socio-Economic Characteristics, subsequently referred to by its Spanish acronym ENCASEH) for all households living in the rural isolated communities which are eligible to receive Progresa. Up to the year 2000, this corresponded to about 3 million household interviews. The data collected include information on educational attainment, monetary income, durable goods, labor force participation as well as indigenous status. To facilitate analysis of the characteristics of this population, Progresa carried out a random (representative) sample equal to about 120,000 households. This data, which provides cross-sectional information for all 32 Mexican states and is the main data source for this study, 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 (estimated at 60 percent) of all indigenous households in Mexico.

Nevertheless, to insure that the results are valid in making inferences about the Mexican population,[8] a nationally representative survey is also used. A separate national sample of the same ENCASEH questionnaire was carried out in 1997 and provides a convenient way to compare the results. This national sample includes 9,910 households and is representative of both urban and rural areas in Mexico.[9]

The dependent variables address two short and long term educational outcomes for boys and girls between the ages of 6 and 18: enrollment and years of completed schooling, respectively.[10] Finally, household and student level data are supplemented with school level information from the Secretary of Public Education (SEP), which allows us to link the characteristics of available schools to children's educational outcomes. This data, drawn from a census collected by the SEP at the time the household data were surveyed, contains specific information about each school, including number of pupils, education of teachers and characteristics of the school infrastructure. In particular, at the primary level, information is used

[8] For instance, the results could be biased by sample selection if indigenous individuals in marginalized areas are not representative of all indigenous individuals.

[9] To avoid confusion, this survey will be referred to as the national ENCASEH, whereas the ENCASEH drawn from the rural communities will be called the 32-state ENCASEH.

[10] The analysis was also performed using the schooling gap measure, defined as age-years of schooling –6: 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

on two parallel programs offered by the SEP (regular and bilingual) to analyze the effect of bilingual education on the educational achievement of indigenous children.

According to the 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 thus clear that this 32-state sample has a much larger concentration of indigenous families than the national level.

5. Methodology and Results

Disentangling Economic Conditions

The previous descriptive non-parametric results clearly showed that indigenous monolingual children lagged behind in schooling outcomes. However, this descriptive evidence cannot determine whether the poor school performance of indigenous children simply reflects cultural and language barriers or in fact represents the inferior social and economic conditions that indigenous households may face. Disentangling these effects is crucial for policymaking: 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.

The attempt to disentangle the effect of a language barrier from that of socioeconomic resources begins with a regression analysis. First, the association between schooling opportunities—as measured by enrollment and years of completed schooling—and belonging to an indigenous group is assessed. The following relationship is estimated for each household child in the sample:

$$(1)\ S_{ic} = \mathrm{B}_0 + X_{1ic}\mathrm{B}_1 + X_{2ic}\mathrm{B}_2 + \delta_1 INDIG_{ic} + u_c + \varepsilon_{ic}$$

---

obtained with years of schooling, so the former is not reported. These results showed that indigenous children have, 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's and father's education, and age, and measures of household wealth and dwelling characteristics. These measures include: ownership of land, access to water and electricity, whether the floor of the house is made of cement and ownership of durable goods including a refrigerator and a stove. INDIG is an indicator of whether the child belongs to an indigenous group, which takes the value of one if the child speaks an indigenous language. The model also includes a community fixed effects u_c, that sweeps out any community time-invariant characteristic which may be correlated with indigenous child schooling outcomes, such as local infrastructure or cultural effects. ε_{ic} corresponds to an error component that reflects all remaining unobserved characteristics of the model.

The particular hypothesis of interest for this study is whether children belonging to an indigenous group display lower school productivity, all other things being equal. In other words, the hypothesis being tested is whether δ is different than zero. A negative coefficient would imply a negative effect of group membership with respect to school opportunities, suggesting that indigenous children are in a disadvantaged position relative to their non-indigenous classmates.

Table 2 shows the determinants of years of schooling and in particular, the effect of being indigenous. Here, we begin with a general measure of indigeneity including indigenous children who are either bilingual or monolingual. The analysis will later disaggregate the two groups.

Table 2 reports a number of specifications, beginning with a minimal specification in which years of completed schooling only depend on child characteristics. Subsequently added specifications include parental characteristics, household wealth indicators and community effects. This makes it possible 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

on average, a schooling gap 2.3 percent higher than non-indigenous children.

effect is contaminated with a human capital effect (older parents tend to be less educated since they belong to older generations), parental levels of education are furthered controlled for in column (4). The effect of mother's age rises marginally and the father's 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 either 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, controls are included for whether the dwelling has concrete floors, walls and ceilings, for whether the household has access to running water and electricity, whether the household owns agricultural land and whether the household owns the durable goods of a refrigerator and stove. All wealth measures are significant and have the expected sign. However and perhaps more importantly, the comparison of columns (5) and (1) shows that the impact of being indigenous on educational outcomes is reduced by half. That is, controlling for parental and household characteristics demonstrates that at least half of observed differences between the indigenous and the non-indigenous is due to family background.

Moreover, it is very likely that indigenous households live in poorer communities with poorer infrastructure relative to the rest of the population. Therefore, failing to control for community observed and unobserved heterogeneity could cause an overestimation of 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 important to mention that the variation across indigenous groups derives mostly from comparisons across communities. Most rural communities in the sample tend to be either primarily indigenous or non-indigenous, and only about 10 percent of the 26,079 communities with at least one child in the sample have both indigenous and non-indigenous children.[11] Nevertheless, the fact that indigenous coefficients do not change significantly when community controls are added suggests

[11] About 86 percent of the variation with respect to indigenous derives from variation between communities.

not only that the potential sample selection bias is low, but also that family background variables capture most of the community effects.[12]

Column (7) presents the same specification as in column (6), but now a *national* sample is used, though one restricted to rural areas. This resulting sample is representative of all rural areas in Mexico.[13] The estimated effects of indigeneity are about double the size than those reported in the ENCASEH 32-state sample. This is perhaps not surprising as this (national) sample is much more heterogeneous than the sample under study, which only contains households living in highly marginalized communities. However it is also likely that a more heterogeneous sample makes it less successful to control for a household's economic (unobserved) circumstances, and thus part of the estimated effect of indigineity in this national sample reflects uncontrolled economic factors at the household level. The rest of the analysis continues with the poorer and more homogeneous ENCASEH 32-state sample, which better controls for household resources and thus better isolates 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 3 presents specifications (5) and (6) for the second schooling outcome, current school enrollment. For comparability, the last two specifications of Table 2 are presented as the first two columns of Table 3. Columns (3) and (4) correspond to OLS and community fixed effects specifications for the probability of current enrollment, respectively.[14] In general, the results are similar to those using years of completed schooling: when household-level variables and community fixed effects are controlled for, indigenous children remain likely to do worse than their non-indigenous classmates. Column (2) shows that indigenous children are, on average, 3.2 percent less likely to be enrolled in school even after controlling for household and community characteristics.

[12] 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 located in wealthier communities, whereas agriculturally oriented households are more likely to live in rural communities with relatively less development.

[13] Urban areas are excluded 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 the ENCASEH-national sample, only 101 report speaking an indigenous language.

[14] For comparability purposes, and given the large sample size, the analysis exploits the fact that the Linear Probability Model (LPM) coefficients are consistent and estimates the probability of school enrollment using OLS models.

Note that the coefficient of being indigenous on the likelihood of currently attending school reverses in sign after community fixed effects are controlled for. This is not the case for years of schooling, which suggests that community unobserved characteristics are differentially correlated to schooling attendance of indigenous children. The last two columns of Table 3 attempt to clarify these differences. In Columns (3) and (4) the indigenous are divided into two groups, *monolingual* indigenous and *bilingual* indigenous to show the differential impact on school enrollment.[15] 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,[16] show that community unobserved characteristics affect differentially the schooling enrollment of monolingual indigenous children and bilingual indigenous children.[17] Therefore, the regressions presented above are, to some extent, an average effect of these two "types" of children and may mask important differences between the two groups. Furthermore, the persistent negative effect of being indigenous monolingual, as shown in Table 3, columns (3) and (4), suggests 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

Next, in an effort to disentangle the possible cultural or language barrier effect from other factors, the sample is restricted to indigenous children who are indigenous, dividing them into two groups: those who speak Spanish (bilingual) and those who are monolingual. The purpose 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. The following specification is used:

$$(2)\ S_{ic} = \mathrm{B}_0 + X_{1ic}\mathrm{B}_1 + X_{2ic}\mathrm{B}_2 + \delta_2 MONO_{ic} + u_c + \varepsilon_{ic}$$

[15] As in previous specifications, non-indigenous is the omitted child category.
[16] As discussed further below, the language division is considered to be endogenous to the determinants of educational outcomes.
[17] A more detailed examination of the characteristics of the communities reveals 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 obviously endogenous to the model, however, and this specification is therefore not presented in the regressions.

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 $\varepsilon_{i\,c}$ corresponds to all remaining unobserved characteristics.

Table 4 presents the findings on the impact of only speaking a native language on years of completed schooling and school enrollment of 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 be currently enrolled in school (Column 4). These results make clear the large educational disadvantage that results when indigenous children do not learn Spanish relative to their indigenous counterparts who do. 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 the size of the coefficient when whether the child is monolingual is controlled for. 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 language has no significant effect on the child's human capital assessment after the child's 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.

It is important to emphasize here, however, that the variable *MONO* is unlikely to be exogenous to schooling determinants. 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 *being 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. The 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.

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, the language spoken by the child is treated as endogenous, and the child's probability of being monolingual is instrumented with his/her parents' ability to speak Spanish as a second language.[18]

The results indicate that the mother's and father's ability (inability) to speak Spanish is a good instrument for a child's ability (inability) to speak Spanish. First, the mother and father's ability to speak Spanish does not significantly affect their child's schooling outcomes once a child's own ability to speak Spanish is controlled for. Second, the child's probability of speaking Spanish is highly correlated with her/his parents' language ability.[19] Finally, the Basmann (1960) IV over-identification tests show that parental language ability is a good instrument for identifying the structural model.

Columns (5) and (7) in Table 4 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. For both indicators, the effect remains negative and significant, although it

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

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: 2SLS estimates show that a child who 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 4, columns (7) and (6), respectively.) Moreover, Hausman specification tests reject the exogeneity of the child's being monolingual as an explanatory variable.

One final exercise is to examine how the language barrier may change or accumulate over time. The previous descriptive results have suggested a widening difference by age in years of completed schooling between monolingual and bilingual as well as non-indigenous children. The last specification is thus repeated in Table 4 above, using an instrumental variable approach to look at the determinants of years of completed schooling and school enrollment by age group. Table 5 presents the summary of these results which show that initially (for the age group of 6 to 8) the effect of being monolingual on the years of completed schooling is relatively small for indigenous children, showing a reduction of only about 0.1 of one year. The effects of this language barrier, however, greatly increase with age. 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 at early ages (perhaps partially reflecting late school entry of monolingual relative to bilingual children). This negative effect continues to be strong on different age groups with the exception of the oldest group (15 to 18) where the estimated effect of being monolingual is no longer a significant determinant of school enrollment.[20]

In summary, failing to control for endogeneity leads to an overestimation of the language barrier effect. Nevertheless, even after controlling for endogeneity, the effect remains quite large, suggesting that language barriers result in a significant disadvantage in terms of overall human capital acquired by indigenous children, a disadvantage which accumulates as children age.

---

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

[20] 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.

Whereas most indigenous children do in fact learn Spanish, those that do not achieve much lower levels of education. It is these children who will most likely experience the long-term consequences of high levels of extreme poverty as adults.[21]

Disentangling the Effect of Language Barriers: Bilingual Education

The previous analysis has made abundantly clear that indigenous children who learn Spanish have much higher achievements in schooling than those who remain monolingual, which we have argued this is a strong indicator of language barriers. To better understand how these language barriers differ from the effect of remaining unobserved cultural factors, the possible role of bilingual education in improving the relative educational performance of indigenous children is examined. Note that monolingual children may be more affected by cultural traditions that may reduce their school attendance, as opposed to bilingual children who are more likely to have assimilated into more western culture.

The language barrier is isolated from possible remaining cultural effects by exploiting the dual educational system provided by the SEP, which consists of *Spanish-type* and *bilingual-type* schools in rural communities. 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 imposition of one language over the other" (Caso, Zavala, Miranda et al., 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 encouraging bilingual teachers who speak a child's indigenous language to pay more attention to and discriminate less against students who do not speak Spanish.

[21] Models carried out with gender interactions reveal a number of interesting gender differences which, for reasons of space, can only briefly be discussed here. Replicating the results in Table 4 for girls and boys, the estimated negative effect of being monolingual on years of schooling is approximately 20 percent larger for girls than for 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 than boys report learning Spanish. 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.

If the large educational difference between monolingual and bilingual children is mainly a language effect, it would be expected, first, that a bilingual school would reduce the negative effect of language between monolingual and bilingual indigenous children in their educational attainment; and second, that other school characteristics would have less or no impact in reducing this gap. Thus the following equation is estimated:

$$(3)\quad S_{ic} = B_0 + X_{1ic}B_1 + X_{2ic}B_2 + \delta_2 MONO_{ic} + \delta_3 BILPRIM_c + \delta_4 MONO_{ic} * BILPRIM_c + u_c + \varepsilon_{ic}$$

Model (3) is an extension of model (2) which interacts the condition that the child is monolingual (as opposed to bilingual), with an indicator variable, BILPRIM, that takes the value of one if the community provides a bilingual primary school.

The difference-in-difference nature of the coefficient δ_4, sweeps out the cultural unobserved heterogeneity between bilingual and monolingual children (to the extent that the impact of the unobserved heterogenity does not vary across school programs) and thus better measures the size of the language barriers. The coefficient indicates whether the language gap in education between monolingual and bilingual children is different for children with a bilingual primary school in their community as opposed to those who do not have access to a bilingual primary school. If the language barrier hypothesis is correct, a positive coefficient would be expected. That is, assuming that bilingual schools reduce the language barrier, one should see a narrowing of the educational disadvantage between bilingual and monolingual indigenous children in communities with bilingual schools.

To test specification (3), the ENCASEH 32-states data are merged with data from the Secretary of Public Education (SEP) covering the same data collection period to determine whether a bilingual primary school is available for children in the community where they live. Availability is initially defined using the school 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, the distance in kilometers is calculated to the nearest community with a school, with a maximum distance of up to 5 km,[22] and the characteristics of a nearby school or schools are used to represent the available supply of

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

bilingual schools for the child. Nevertheless, in the hopes of better capturing behavior in terms of the school that children actually attend, the estimation is restricted in several ways:

First, the empirical model restricts attention to communities where there is a primary school located within and not outside the community. A key underlying assumption in this matching procedure is that primary-level children do not attend community schools other than their own; this would prevent us from matching the *true* underlying community school infrastructure to the corresponding child. It appears, however, that *conditional on the community having a primary school*, children may be unlikely to commute to other communities, given that the sample is characterized by households with very low resources located in very isolated communities.

It may still be the case in the sample, however, that a community may have more than one available primary school. In this case, there would be no way to verify which primary school each child attends and it would be necessary to arbitrarily construct "averages" of school characteristics. Thus, to maximize the probability of matching school infrastructure to each child, the sample is further constrained to communities where there is only one available primary school.[23]

Finally, the analysis is restricted to current schooling enrollment, the short run schooling outcome, to avoid the potential problem of migration. That is, during their life course, some children may have attended schools other than the community school where they live at the time of the interview.

Table 6 summarizes the findings. Columns (1) and (2) suggest that results drawn from our restricted sample are consistent with earlier evidence. IV estimates on monolingual indigenous children attending primary level with only one school in their community suggest they are 11 percent less likely to be currently enrolled relative to bilingual children, as opposed to the unconstrained sample with a probability of 10 percent (see Table 4). Columns (3) to (5) display the results of the difference in difference estimation. Here, the focus is on the difference and difference coefficient (δ_4), which should be unaffected by the possible endogeneity of

[23] Overall, 55 percent of indigenous children in the sample have access to a bilingual school. Within this total, 51.8 percent of bilingual children have access to a bilingual school, whereas the figure for monolingual indigenous children rises to 71.3 percent.

language. From column (3), it can be seen that this coefficient implies a 3.5 percentage point reduction in the language barrier gap; however, it is not significant, given the large sample size.[24]

A potential empirical problem which arises here is that of endogenous program placement (Rosenzweig and Wolpin, 1986). It is likely that the Mexican government locates bilingual schools precisely in areas with higher indigenous populations 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 may be biased downward. To correct for possible endogenous program placement, we further control for observed and unobserved community heterogeneity is further controlled for and model (3) is re-estimated using community fixed effects. As expected, column (4) shows that once community fixed effects are controlled for, the difference-in-difference coefficient increases in magnitude but remains not significant, given the significance levels used.

This may, however, be due to the large heterogeneity in school quality of the Mexican educational system (see López, 2001). In the present sample, teachers in bilingual schools (who are fluent in a native language) demonstrate lower levels of education than their non-indigenous colleagues in other primary schools.[25] It may correctly be assumed that community effects would account for level effects (there is one school per community). Nevertheless, these school quality effects may still bias downward the double difference coefficient in the event of nonlinear spurious correlation between school performance of monolingual children and low quality native schools that are largely attended by monolinguals.

Therefore, the final specifications in Table 6 further control for additional school quality variables, including the education level of teachers, the student teacher ratio, and the interaction of these variables with a dummy variable that indicates whether a child is monolingual.[26] The results are as expected. The double difference estimate on the interaction term of a child being monolingual with access to a bilingual school further increases to 6.8 percentage points and

[24] 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 statistic greater than 3.15 is judged significant in the regressions in the table.

[25] López, 2001 shows that test scores in bilingual primary schools are lower than for any other type of primary school in Mexico.

[26] Teachers are classified into three categories. The first group includes teachers with secondary school or less, whereas the second group includes teachers with a high school education. The last group includes teachers with university education or above, including both those who attended a general university or university specialized in training for teachers. Because of the number of very small schools in our sample, a dummy variable specification is used for each group, e.g., schools with at least one teacher in the relevant group.

becomes significant. Thus, bilingual schools significantly reduce the enrollment gap between monolingual and bilingual children. The estimated reduction is quite large, corresponding to a reduction of over 50 percent in the gap in school attendance, compared with the original estimated IV language barrier of 11 percentage points. Moreover, it is also important to note that in the final specification (column 6), none of the school quality variables interacted with a child's language is significant. An F test of the joint significance of all of the school quality interactions (excluding access to a bilingual school) confirms that these interactions are jointly insignificant. Among school characteristics, only the bilingual educational attribute (teachers who speak the native language, textbooks in the indigenous language, etc.) reduces the enrollment gap between monolingual and bilingual children. These results are interpreted as confirming the existence of a large language barrier responsible for schooling inequality among the indigenous.

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. Educational attainment is likely an important determinant of the level of social exclusion that indigenous children may face in the future. With the aim of analyzing how the education level of the indigenous can be improved, this paper has provided a first step towards a diagnostic of the factors affecting educational attainment of indigenous children.

The paper has shown that indigenous children on average fare worse in educational outcomes than non-indigenous children, even within highly marginalized rural areas of Mexico. Nonetheless, there is great heterogeneity within the indigenous population. When indigenous children learn Spanish, they achieve educational outcomes that are almost equivalent to their non-indigenous counterparts. When they do not learn Spanish, however, their educational outcomes are far inferior. In this way, the level of social exclusion that the indigenous may face is not homogenous. There are important differences between bilingual and monolingual indigenous children.

The analysis has shed some light on the explanatory factors for why some monolingual indigenous children perform worse 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 greatly reduces their educational achievement. Language barriers can thus be taken as an important factor behind the continued social exclusion of the monolingual indigenous.

Bilingual education is a relatively new phenomenon in education in Mexico. The results are suggestive of potential positive effects of bilingual primary schools in the sense that the negative effect of estimated language barriers is reduced when indigenous children have access to bilingual education. This is suggestive that bilingual schools can reduce the degree of social exclusion that monolingual indigenous children face. Nevertheless the impacts and results of bilingual schools constitute an important area for evaluation and should continue to be monitored. It is clearly still too early to speculate on the 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 and thus increase their level of social exclusion as adults.

References

Barro, R., and J. Lee. 1996 "International Measures of Schooling Years and Schooling Quality." *American Economic Review* 86(2): 218-223.

Basmann, R.1960. "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics." *Journal of the American Statistical Association* 55(292): 650-659.

Behrman, J. 2001. "Targeting Social Protection to the Poor: What are the Tradeoffs?" Philadelphia, United States: University of Pennsylvania, Department of Economics. Mimeographed document.

Behrman, J., and A.B. Deolalikar. 1991. "School Repetition, Dropouts, and the Rates of Return to Schooling: the Case of Indonesia." *Oxford Bulletin of Economics and Statistics* 53(4): 467-480.

Behrman, J., and J.C. Knowles. 1999. "Household Income and Child Schooling in Vietnam." *World Bank Economic Review* 13(2): 211-256.

Behrman, J., and P.E. Todd. 1999. "Randomness in the Experimental Samples of PROGRESA (education, health, and nutrition program). Report submitted to PROGRESA. Washington, DC, United States: International Food Policy Research Institute. <http://www.ifpri.org/themes/progresa.htm>

Behrman, J., P. Sengupta and P. Todd. 2000. Final Report: The Impact of PROGRESA on Achievement Test Scores in the First Year. Washington, DC: International Food Policy Research Institute. <http://www.ifpri.org/themes/progresa.htm>

Behrman, J., N. Birdsall and M. Székeley. 1999. "Intergenerational Mobility in Latin America: Deeper Markets and Better Schools Make a Difference." In: N. Birdsall and C. Graham, editors. *New Markets, New Opportunities? Economic and Social Mobility in a Changing World*. Washington, DC, United States: Brookings Institution/Carnegie Endowment for International Peace.

Blake, J. 1989. *Family Size and Achievement*. Berkeley, United States: University of California Press.

Caso, A., S. Zavala, J. Miranda *et al*. 1981. *La Politica Indigenista en Mexico: Métodos y resultados*. Mexico, D.F., Mexico: Consejo Nacional para la Cultura y las Artes.

Chiswick, B.R. 1991. "Speaking, Reading, and Earnings among Low-Skilled Immigrants." *Journal of Labor Economics* 9(2): 149-170.

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

Church, J., and I. King. 1993. "Bilingualism and Network Externalities." *Canadian Journal of Economics* 26(2): 337-345.

Duignan, P. 2000. "Bilingual Education: A Critique." Stanford, United States: Hoover Institution. Mimeographed document.

Heer, D.M. 1985. "Effects of Sibling Number on Child Outcome." In: R.H Turner and J.F. Short, Jr., editors. *Annual Review of Sociology* 11: 27-47. Palo Alto, United States: Reviews, Inc.

Hausman, J.A. 1978. "Specification Tests in Econometrics." *Econometrica* 46(5): 1251-1272.

Hauser, R.M. and D. Kuo Hsiang-Hui. 1995. "Does the Gender Composition of Sibships Affect Educational Attainment?" CDE Working Paper 95-106. Madison, United States: Center for Demography and Ecology.

Hernsein, R.J. and C.Murray, C. 1994. *The Bell Curve*. New York, United States: Free Press.

Instituto Nacional de Estadística Geografía e Informática of Mexico (INEGI). 1994. "Estadísticas Históricas de México." Mexico: INEGI.

INEGI. 2000. "Información preliminar, Censo Nacional de Población 2000." Mexico, D.F., Mexico: INEGI.

John, A., and K. Yi. 1997. "Language, Learning and Location." Charlottesville, United States: University of Virginia. Mimeographed document.

King, E., and L.A. Lillard. 1987. "Education Policy and Schooling Attainment in Malaysia and the Philippines." *Economics of Education Review* 6(2): 167-181.

Lang, K. 1986. "A Language Theory of Discrimination." *Quarterly Journal of Economics* 101(2): 363-382.

Lazear, E. P. 1999. "Culture and Language." *Journal of Political Economy* 107(6): 95-126.

Lombardo Toledano, V. 1991. *Escritos Acerca de la Situacion de los Indigenas*. D.F., Mexico: Centro de Estudios Filosoficos, Politicos y Sociales Vicente Lombardo Toledano.

López Acevedo, G. 1999. "Learning Outcomes and School Cost-Effectiveness in Mexico." Policy Research Working Paper 2128. Washington, DC, United States: World Bank.

López Acevedo, G. 2001. "Professional Development and Incentives for Teacher Performance in Schools in Mexico." LCSHD Paper Series 64. Washington, DC, United States: World Bank.

Modiano, N. 1988. "Public Bilingual Education in Mexico." In: C.B. Paulston, editor. *Handbook of Bilingualism and Bilingual Education.* New York, United States: Greenwoood Press.

O'Faircheallaigh, C. 1998. "Resource Development and Inequality in Indigenous Societies." *World Development* 26(3): 381-394.

Palafox, J.C., J. Prawda and E. Velez. 1994. "Primary School Quality in Mexico." *Comparative Education Review* 38(2): 167-180.

Panagides, A. 1994. "Mexico." In: G. Psacharopoulos and H.A. Patrinos, editors. *Indigenous People and Poverty in Latin America: An Empirical Analysis*. Washington, DC, United States: World Bank.

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

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

Rosenthal, A.S., K. Baker and A. Ginsburg. 1983. "The Effect of Language Background on Achievement Level and Learning Among Elementary School Students." *Sociology of Education* 56: 157-169.

Rosenzweig, M., and K.I. Wolpin. 1986. "Evaluating the Effects of Optimally Distributed Public Programs: Child Health and Family Planning Interventions." *American Economic Review* 76(3): 70-82.

Schultz, T.P. 2000. "School Subsidies for the Poor: Evaluating a Mexican Strategy for Reducing Poverty." New Haven, United States: Yale University. Mimeographed document.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461- 464.

Scott, J. 2000. "Who Benefits from the State in High-Inequality, Middle-Income Countries? The Case of Mexico." Mexico, D.F., Mexico: Centro de Investigación y Docencias Económicas. Mimeographed document.

Secretaria de Educación Pública. 2000. *Perfil de la Educación en México*. Mexico, D.F., Mexico: Secretaria de Eduacación Pública.

Schwartz, G. 1978. “Estimating the Dimension of a Model. *Annals of Statistics* 6: 461-64.

Terborgh, A., J.E. Rosen, R. Santiso Galvez *et al*. 1995. “Family Planning among Indigenous Populations in Latin America.” *International Family Planning Perspectives* 21(4): 143-149.

Table 1.
Indigenous Status, Education and Labor Income: Adult Men Aged 16 to 55

	Indigenous	Non-indigenous	T stat-sig. Dif.
National sample	967.9	1748.4	9.8
	(1138.4)	(1829.9)	
By years of completed schooling			
0	570.0	1010.3	3.7
	(750.9)	(1118.9)	
1 to 5	695.0	1096.2	5.4
	(586.5)	(1010.8)	
6	830.9	1324.5	4.2
	(666.1)	(1298.6)	
7 to 9	1339.9	1520.0	1.0
	(1720.1)	(1396.7)	
More than 9	2359.7	2835.1	1.5
	(1702.8)	(2461.5)	

Figure 1.

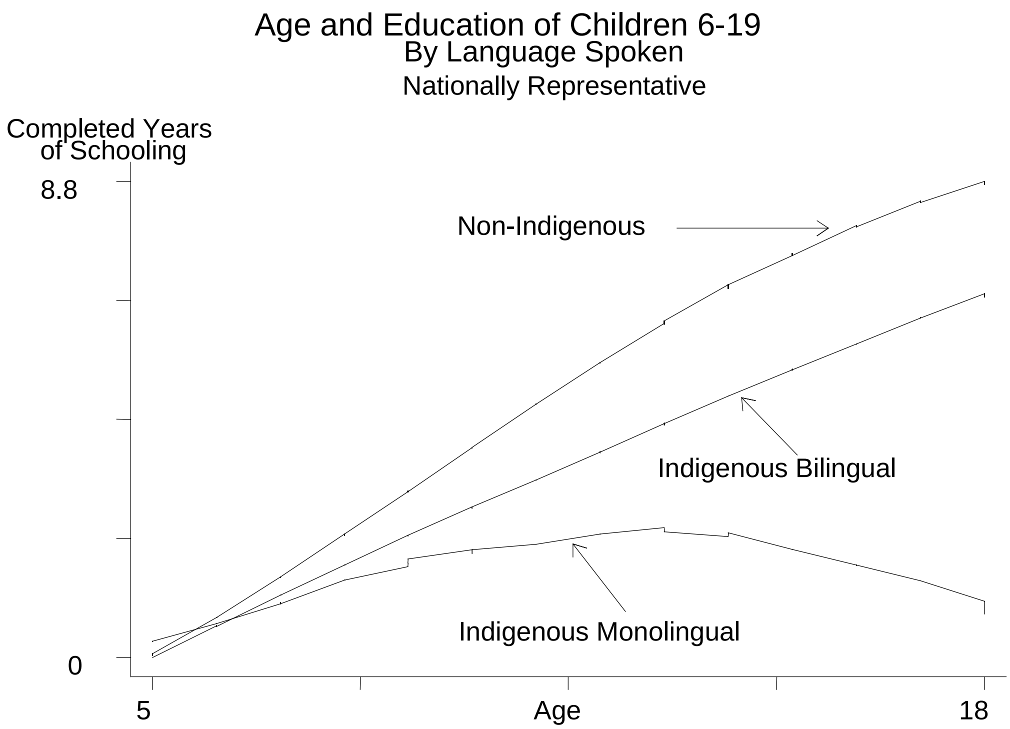

Figure 2.

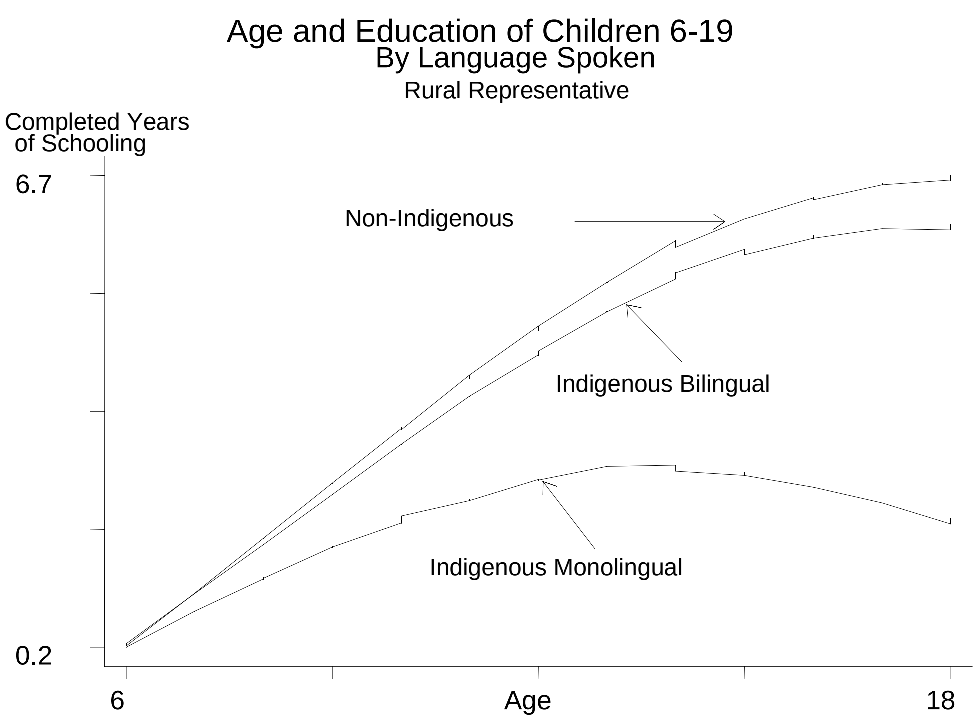

Figure 3.

Probability of Indigenous Monolingual
By Schooling Gap

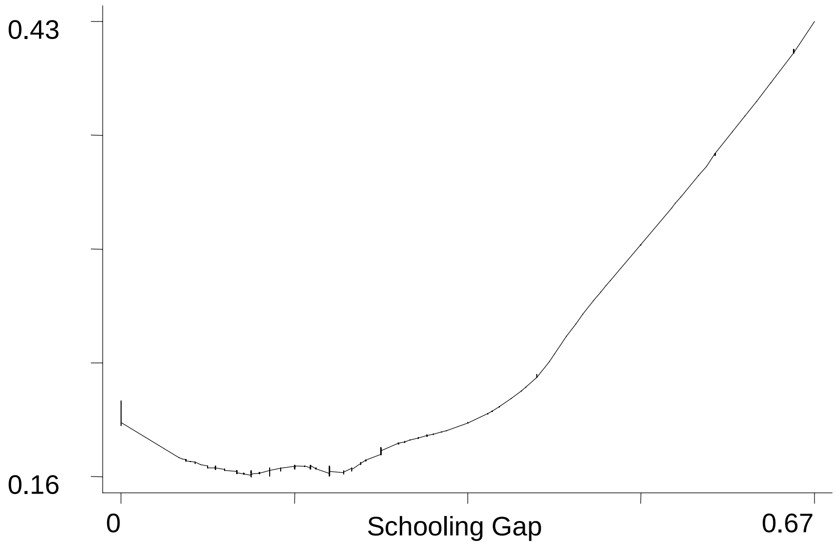

Table 2.
Determinants of Years of Completed schooling: Effect of being Indigenous, Children Aged 6 to 18

	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

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

Table 4.
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

Table 5.

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	-0.414	-1.602	-2.265
	[0.030]**	[0.071]***	[0.132]***	[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	-9.364	-15.181	-4.268
	[1.555]***	[1.712]***	[2.967]***	[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 6.
School Enrollment: Language and Bilingual School Interactions

	OLS	IV	OLS	OLS	CFE
	[1]	[2]	[3]	[4]	[5]
Child only speaks indigenous language	-15.132	-11.443	-18.523	-16.38	-12.981
	[0.653]***	[1.426]***	[1.333]***	[3.909]***	[4.645]***
Bilingual school			2.83	2.474	
			[0.508]***	[0.652]***	
Indigenous language * bilingual school			3.622	6.178	6.793
			[1.512]**	[1.772]***	[2.000]***
Indigenous language* teacher education level 2				-4.251	-4.16
				[1.512]***	[1.792]**
Indigenous language * teacher education level 3				0.56	-6.235
				[2.225]	[2.734]**
Indigenous language * student teacher ratio				-0.057	0.033
				[0.096]	[0.104]
F(3, 21268)				3.25	3.00
Observations	21274	21274	21274	21268	21268
R-squared	0.38	0.38	0.38	0.38	0.38
Number of communities					3266

Notes: Results using Encaseh 32-states sample for indigenous children between 6 to 12 years old in communities with only one primary s
Robust standard errors in [brackets]. Coefficients marked with (***) are significant under Schwartz' (1978) with a t statistic no smaller tha
models include child characteristics: sex and age groups; parental characteristics: father's age, mother's age, father and mother education g
household assets: cement floor, water and electricity, agricultural land owning, refrigerator and stove. Columns (4) and (5) also include scho
characteristics: teachers educational categorical variable levels, student to teacher ratio and number of classrooms.

Appendix

Table 1A.
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. Descriptive Statistics: 32 State Sample by Ethnic status

Variable	Non-Indigenous	Indigenous	
		Bilingual	Monolingual
	[Std. Dev.]	[Std. Dev.]	[Std. Dev.]
Age	11.739	12.070	10.116
	[3.685]	[3.626]	[3.579]
Gender (Boy=1)	0.507	0.513	0.467
	[0.499]	[0.500]	[0.499]
School enrollment (percentage)	69.587	69.195	61.035
	[46.003]	[46.169]	[48.770]
Years of schooling	4.081	3.925	1.576
	[2.872]	[2.702]	[1.873]
Household size	6.964	7.135	7.296
	[2.425]	[2.386]	[2.373]
Father's education	3.362	2.992	2.224
	[2.702]	[2.613]	[2.274]
Mother's education	3.142	2.046	1.133
	[2.682]	[2.419]	[1.866]
Father's age	42.750	42.885	40.977
	[8.697]	[8.700]	[9.072]
Mother's age	38.252	38.578	36.657
	[0.082]	[7.893]	[8.118]
Father only speaks indigenous language	0.000	0.055	0.287
	[0.000]	[0.227]	[0.452]
Mother only speaks indigenous language	0.000	0.176	0.665
	[0.000]	[0.381]	[0.472]
Cement floor	0.143	0.058	0.023
	[0.350]	[0.234]	[0.151]
Hhold has water and electricity	0.140	0.085	0.061
	[0.347]	[0.279]	[0.239]
Hhold owns agric. land	0.512	0.719	0.795
	[0.500]	[0.449]	[0.404]
Hhold has refrig. and stove	0.227	0.036	0.009
	[0.419]	[0.186]	[0.092]
Observations	170,495	42,640	9,561