Mandatory Schooling and the Language-based Educational Gap

Lyliana Elizabeth Gayoso  
Ph.D. Candidate  
lyliana.gayosogomez@email.wsu.edu  
School of Economic Sciences  
Hulbert Hall 323 E, Washington State University, Pullman WA 99164-6210

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE

Washington State University  
School of Economics Sciences  
Seminar Series 2014

Copyright 2014 by Lyliana Gayoso. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Mandatory Schooling and the Language-based Educational Gap

Lyliana Gayoso*

Abstract

Educational gaps due to differences in language exist in several countries. Different policies have been pursued to enhance educational levels in these countries without much success. In Paraguay, a new mandatory and free basic education policy was undertaken in 1994. I examine the effects of this policy on educational attainment of individuals from different linguistic backgrounds to assess whether this type of policy can help to reduce the existing language-based educational gap. The Difference-in-Differences estimates suggest a shift in educational attainment of all individuals regardless of their language. The implications of these results on the language-based educational gap are important. These findings suggest that the language-based educational gap persists and so does the need to reduce this gap.

JEL classification: I21, I24, O20

Keywords: education, educational gap, language

* School of Economic Sciences, Washington State University, Pullman, WA, USA.
e-mail: lyliana.gayosgomez@email.wsu.edu
1. Introduction

Raising educational attainment has long been a focus of policy makers in developing countries. Most of the implemented educational policies have been based on the idea that education positively impacts economic growth and development. Furthermore, it is believed that enhancing educational attainment leads to high social welfare gains, in particular when individuals from different backgrounds or those at the bottom of the education ladders increase their schooling attainment (Lillard & Willis 1994, Patrinos & Psacharopoulos 1995). However, despite policy makers' best efforts, developing countries continue to be characterized by low levels of education. Around 200 million young people in the world still lack basic literacy skills (UNESCO 2012).

To increase educational attainment, policy makers and researchers have focused on enhancing access to education, in order to reduce existing educational gaps. Such gaps emerge due to differences in socioeconomic status (Garcia Palomer & Paredes 2010), as well as gender (Banerjee et al. 2000; Kingdon 2002, 2005). Yet another gap, the language-based educational gap, has been observed by several researchers (Chiswick et al. 2000; Glewee & Kremer 2006), but has not been given as much attention. Educational attainment gaps associated with language differences emerge in numerous countries where language heterogeneity exists within the population. According to Glewee and Kremer (2006), language heterogeneity, differences in educational background, and schooling quality can contribute to a mismatch between the curriculum and the needs of the typical student. The adoption of a single curriculum in developing countries is often designed to meet the needs of elite students, at the expense of the most disadvantaged children. This leads to educational attainment disparities by language, which in turn results in large gaps in labor market outcomes.

This study evaluates the success of a mandatory schooling policy that seeks to increase the average level of educational attainment in the population by promoting equity in the education system of Paraguay (Rivarola 2000; World Bank 2010). This policy, implemented in 1994, expanded mandatory free basic education from six to nine years to increase years of education. Even though high returns of education in developing countries have been found (Hanushek 1995; Duflo 2001), the investment on education remains low. The underlying reasons remain an open question in the literature.

In particular, this seems to be often a case in countries with indigenous languages (Patrinos et al., 1994; Chiswick et al., 2000), where the immigrant population is large (Colding, 2006), or in countries with linguistic heterogeneity.
schooling of disadvantaged students. These students are often those who do not speak the dominant language or the language of instruction. Paraguay is a country where two official languages are present: Spanish and Guaraní. The levels of education between Spanish and Guaraní speakers have differed substantially historically\(^3\), with Guaraní speakers having much lower levels of education. The World Bank (2010) report indicates that in Paraguay, those with low educational attainment and with the most limited access to education are frequently the most impoverish. Therefore, while this policy was universally implemented, I consider the hypothesis that those who were most affected are the language-disadvantaged students. Furthermore, these individuals reside mostly in rural areas and are Guaraní speakers. Demand for schooling of these students in the absence of the policy would have been less than the years enforced by the mandatory educational policy\(^4\). Because these students often quit schooling at early stages, their schooling levels are most likely to be affected by the policy\(^5\). The effects on educational attainment could further translate into an impact on labor market outcomes. In contrast, Spanish speakers have on average the highest levels of education, which are beyond compulsory schooling years. Therefore, Spanish speakers are less likely to be affected by this policy. Hence, an empirical implication of the hypothesis stated in this study is that Guaraní speakers, whose average educational attainment is lower than the mandatory school years, will increase their schooling relative to Spanish speakers. If there is evidence in favor of this hypothesis, then enhancing educational attainment of language-disadvantaged students may reduce the language-based educational.

I empirically examine this hypothesis by using repeated cross-sectional data for Paraguay from the period between 2002 through 2012. This dataset, unlike others, contains exact information of date of birth for each individual. To define my identification strategy I consider the following. First, the new policy was implemented for all students entering the first grade of basic education in 1994, and all successive cohorts. Second, the admission in the first grade requires the student to be six years of age upon entrance and to turn seven years old by the month

---

\(^3\) In an early study, Patrinos et al. (1994) stated that “language selectivity exists at all levels of education and income in Paraguay”, thus providing evidence of the presence of language-based educational gaps.

\(^4\) I abstain to argue that “optimal” years of schooling would have been lower since in developing countries, marginal benefits and marginal costs of education are not equated. Lillard & Willis (1994) indicate that according to the Becker-Tomes model, poor families do not demand optimal years of schooling, but a lower amount.

\(^5\) This is based on the literature of compulsory schooling laws. This literature has found that compulsory schooling laws increase educational attainment by compelling students to attend school longer. This higher level of education, in turn, results in higher levels of wages (Angrist and Krueger 1991; Margo and Finegan 1996; Oreopolous 2006 a,b ).

4
of August of each academic year (academic year runs from mid February to early December). Hence, the research design uses exact date of birth and academic year at entry to identify the cohorts of individuals that were and were not exposed to this new policy. The evaluation of this type of intervention is difficult. The effects of the policy may be confounded with other structural and/or macroeconomic effects that may contribute to educational attainment increasing over time in Paraguay (World Bank 2010). Isolating the effects of the policy becomes a difficult task. The Difference-in-Differences (DID) econometric technique is used, with the first difference on the exposure to the new policy, and the second difference on language using Spanish speakers as the reference group. Therefore, the DID technique allows me to use exact date of birth, academic year at entry, and differences by language (variables exogenous to the new policy), to isolate the effects of the policy.

The analysis focuses on outcomes of individuals of age 22, since they are more likely to have reached their desired educational attainment. I find evidence of an upward shift in the range of 0.24 and 0.72 years of schooling for Guaraní speakers relative to Spanish speakers, however this is not statistically significant at conventional levels. However, this coefficient is not statistically at conventional levels. This implies that the effects of the policy are not as large for Guaraní speakers as hypothesized. Similar results were found for bilingual speakers. I provide an explanation for these results. Guaraní speakers are frequently in the bottom of the income distribution, and their parents have low levels of educational attainment. This could be evidence of the persistence of intergenerational transmission of education and poverty. Also they are highly concentrated in rural areas of residency, where access to education might be limited. In addition to this, while they are monolingual in Guaraní, the language of instruction in Paraguay is primarily in Spanish. This reveals two important points. First, Guaraní speakers might experience difficulty learning in a language that was completely unknown for them. Second, it is likely that in rural areas teachers’ proficiency in Spanish is not adequate to instruct children in that language or even to comprehend schooling materials. This is further compounded by the challenging facing those living in poverty.

Finally, while empirical work has examined the effects of mandatory schooling policies on educational gaps focusing mainly on gender gaps (see for instance Spohr, 2003), little is known about the effects of educational policies on language-based educational gaps. Best
practices in education initiatives in developing countries are far from being established, and several studies have recognized the difficulty of increasing levels of education of individuals (Spohr 2003; Glewwe & Kremer 2006). Hence, this study extends the existing literature by examining the effects of a mandatory schooling policy on language educational-based gaps. In addition, this paper has an empirical contribution to the literature on the returns to schooling in developing countries. My findings suggest that an increase in mandatory schooling may not be the most effective way to enhance schooling of language-disadvantaged students and consequently reduce the language-based educational gap.

The rest of the paper is organized as follows. Section 2 presents a brief background of the reform and the education system in Paraguay. Section 3 outlines the empirical framework, providing a description of the identification strategy and the econometric model. Section 4 present descriptive statistics; while Section 5 reports the results obtained. Finally, Section 6 discusses the findings of the research and provides final remarks.

2. The Educational System in Paraguay

In 1994, the strong desire to increase the education levels of Paraguayans, in the hope that such increases would translate into an enhancement of poverty levels,\(^6\) led to the implementation of a new educational policy. This new policy involved the expansion of compulsory years of schooling from 6 to 9 years. This expansion entailed a new structure of the basic education level, which incorporated one more level. Previous to the policy, basic education was composed of two levels. The first level included grades 1 to 3, and the second level, grades 4 to 6. With the implementation of the policy, one level was added to these two, becoming the third level. This third level included three grades, from grades 7 to 9. This change in the organization of the basic education was undertaken with the objective of expanding mandatory free basic education to all the nine grades. The organizational change and the expansion of the mandatory and free years of education were accompanied by the distribution of free books in public schools (for the first time), an increase in hours of schooling, and the reform of the national curriculum (Rivarola, 2000; World Bank, 2008)\(^7\).

---

\(^6\) The IDB 2004 loan proposal report clearly states that education is a key factor for poverty reduction in Paraguay.

\(^7\) For further details see Rivarola 2000.
The new policy was launched in 1994 and affected all students entering the first grade of basic education in the same year. It was gradually implemented one grade per year, so that by 2003, all the school grades of the basic education system were affected by the reform. In addition to this, in 1998, when the first cohort under the new system was in 5th grade, the government decided to include the previous cohort in the new system. Thus, these students were affected by this new policy beginning in 1998 onwards.8

A. Language heterogeneity and the new mandatory educational policy

Paraguay is known for having two official languages: Spanish and Guaraní.9 While Guaraní is an indigenous language, however, its use is highly spread regardless of ethnicity. Interestingly, Spanish is spoken primarily in urban areas, while Guaraní is primarily spoken in rural areas (Gynan 1998). Nevertheless, a large population speaks both languages.

In early nineties, an analysis of the sociological characteristics of the population highlighted an important feature on language usage: a large percentage of the population was monolingual in Guaraní. According to the National Census of 1992, about 50% of the population spoke both Spanish and Guaraní, about 37% was monolingual in Guaraní, only around 7% was monolingual in Spanish, and 6% of the population spoke other languages.

The evaluation of the educational system that preceded the new educational policy revealed high inequalities in access to education, and substantial differences in educational attainment associated with differences in language. In addition to this, it was brought to the attention of policy makers that monolingual Guaraní children were facing significant barriers to progress in school. These barriers were associated to the fact that Spanish had for long been the language of instruction (World Bank 2008). These features on language usage and education provide evidence on the long existence of the language-based educational gap in Paraguay.

In general, the execution of this new educational policy brought significant improvements in terms of enrollment, dropout rates, and repetition rates, enhancing the country’s education standing in the region (World Bank 2008; 2010). However, these results are shown from

---

8 Considering that this cohort was partially affected, it is left out of from the main analysis.
9 Paraguay is a country located in South America. As an earlier Spanish colony, the use and acquisition of the Spanish language came from the conquerors. The Guaraní language, however, is the native language, inherited from the early indigenous people.
10 As part of the evaluation of the educational system, the Institute of International Development (IID) at the University of Harvard and the Paraguayan Center for Sociological Studies (CPES) provided a series of studies, that revealed the major needs of the system.
qualitative evaluations based on the evolution of rates, rather than a quantitative analysis. Therefore, this study contributes to knowledge by providing an exhaustive quantitative analysis of the topic.

Primarily, this paper seeks to understand the effects of a mandatory education policy on educational attainment of individuals with different linguistic background. In particular, it is of primary interest to answer the following question: Can an educational reform based on expanding mandatory free education reduce the language-based educational gap? The rest of the paper is devoted to examine this topic in depth.

3. Empirical Framework

The extant literature that assesses public policies has widely agreed that an ideal evaluation strategy would involve a random assignment of individuals to a treatment (Angrist and Lavy 1997; Angrist and Pischke 2008; Banerjee 2007). In this sense, randomized evaluations of public policies have become popular to implement in developing countries (Glewwe and Kremer 2006). However, in practice, it is very difficult to randomly assign individuals to a treatment or intervention, and randomized evaluations, although ideal, are still costly and difficult to implement in several countries.

In the case of the educational policy analyzed in this study, it is possible to argue that the individuals were not randomly assigned to this policy intervention, since it was implemented nationwide concurrently. Therefore, the evaluation of this type of intervention presents some difficulties. In particular, the nationwide implementation requires the analysis to rely on comparison of cohorts that were and were not affected by this policy (or to compare outcomes pre and post reform). Then, the effects of the policy may be confounded, for instance, with other structural and macroeconomic effects. To overcome these difficulties, this study appeals to Differences-in-Differences econometric technique\(^\text{11}\).

The simultaneous nationwide implementation of this educational policy intervention to the incoming cohorts starting in 1994, provides the opportunity to evaluate the effects of this policy by using the Difference-in-Differences approach. In the set up of this model, outcomes are

\(^{11}\) Some other alternative econometric techniques in this area includes local average treatment effect, and regression discontinuity. However, due to the small sample size the use of these techniques are infeasible. For a more detailed discussion of these and other techniques refer to Angrist and Pischke (2008).
observed for the control and treatment groups, and for the different linguistic groups: Guaraní, Spanish, and bilinguals speakers. Taking into account that the Spanish speakers are the most advantaged linguistic group, the idea behind the DID approach is to compare the average gains in the outcome variable for Guaraní and bilingual speakers relative to Spanish speakers. This implies that the average gains in the control group are subtracted from the average gains in the treatment group for each language group. Furthermore, the average gains obtained by comparing treatment and control groups for Guaraní speakers is subtracted from the average gains obtained for Spanish speakers. Similarly, the average gains for bilingual speakers are subtracted from the average gains for Spanish speakers.

According to Wooldridge (2007), this step allows to remove any biases that could arise due to permanent differences between control and treatment groups. A key assumption in this econometric technique is that in the absence of the new policy, pre-reform trends in the outcome variable would have remained the same, conditional on observed characteristics. Under this assumption, it is possible to evaluate the effects of the policy by comparing average outcomes for the two groups defined.

A. Identification Strategy

The date of birth and academic schooling year jointly determine an individual’s exposure to the new policy. In addition, a second source of identification, the language variation, is considered. To identify the different linguistic backgrounds of individuals, I use language spoken at home as a proxy. The rationale behind this approach is that early education acquired at home is obtained in the language spoken by parents. Therefore, the language used at home is a good proxy of language skills.

Considering that the new mandatory educational policy was implemented on 1994 onwards, it provides the opportunity to examine the effects of this policy on educational attainment within a Difference-in-Differences framework. Within this framework, I compare cohorts of individuals that were not affected by the new policy to the cohorts affected by it. The
first group of cohorts is then defined as the "control group", while the second group of cohorts is defined as the "treatment group".\textsuperscript{12}

Previous to the introduction of the new policy, children in Paraguay were required to attend school between ages 7 and 12, consisting of grades 1 to 6\textsuperscript{13}. With the introduction of the new mandatory schooling policy in 1994, children attended school between ages 6 and 11. However, in practice before and after the policy, admission in the first grade requires the student to be six years at entrance and to turn seven years by the month of August of each academic year. Therefore, all children born from September of 1986 and onwards were affected by this new policy. In contrast, children born before September of 1985 were not exposed to the new policy\textsuperscript{14}. 

The treatment dummy in my model is then defined to take a value of one if the individual in the sample belongs to a cohort that was affected by the new policy and a zero if the individual was born in a cohort that was not exposed to the new educational policy. Furthermore, considering the amount of information available, the main analysis includes five cohorts in the control group and five cohorts in the treatment group (see Table 1 of the Appendix for further details).

\textbf{B. Econometric Model}

The identification strategy can be illustrated within a regression framework. First, consider the difference between the average years of schooling of a cohort that was exposed to the new policy and that of a cohort that was not exposed to the new policy. Second, consider the second difference between the average years of schooling between Guaraní and Spanish speakers, and between bilingual and Spanish speakers. Then, if the mandatory educational policy led to a higher increase in years of schooling of Guaraní and bilinguals speakers relative to Spanish speakers, the difference will be positively related to the implementation of the policy. Considering the observation of repeated cross sectional data, this suggests that the model can be written as:

\textsuperscript{12} The cohort that was partially treated is omitted from the main analysis; however, it is used as part of a robustness check, which will be discussed later in the study.
\textsuperscript{13} For an overview of the educational sector see World Bank 1995.
\textsuperscript{14} The cohort of children born between September of 1985 and August of 1986 is left out from the main analysis, since they started primary school under the old policy and were incorporated to the new policy in 1998, when they entered 6th grade.
\[ S_{it} = \beta_0 + \beta_1 G_i + \beta_2 B_i + \delta_1 d_i + \delta_1 (d_i G_i) + \delta_2 (d_i B_i) + \gamma_1 X_{it} G_i + \gamma_2 X_{it} B_i + \gamma_3 X_{it} S_i + \epsilon_{it} \]

\( i \in \Lambda, t = 2002, \ldots, 2012 \) and where \( S_{it} \) is years of schooling of individual \( i \) in time \( t \), \( d_i \) is a dummy variable that indicates treatment status, or more specifically, indicates whether an individual was exposed or not by the new educational policy; \( G_i \) and \( B_i \) are dummy variables that captures differences in language spoken; \( G_i \) is a dummy variable that takes a value of one if an individual is a Guaraní speaker, while \( B_i \) is a dummy variable indicating that an individual is bilingual speaker. Moreover, these two dummy variables intend to capture possible differences between treatment and control groups within the language group, prior to the policy change; \( X_{it} \) represents a vector of covariates that are time-varying. Finally, \( \epsilon_{it} \) represents the error term of the equation.

The coefficient of interest \( \delta_1 \) represents the change in the outcome of interest, in this case years of schooling, after the implementation of the new mandatory schooling policy for Guaraní speakers relative to Spanish speakers. This is Difference-in-Differences estimator. Similarly, \( \delta_2 \) represents the change in the outcome of interest after the implementation of the mandatory schooling policy for bilingual speakers relative to Spanish speakers.

Caution should be exercised when interpreting the coefficient of interest, \( \delta_1 \) causally. The identification assumption underlying this econometric technique is that there are no omitted time varying effects correlated with the new policy. This identification assumption will be violated if there were other public policies correlated with outcomes that were undertaken at the same time of the mandatory schooling policy. Furthermore, if trends in the outcomes would have differed by date of birth and/or language for unobserved motives even in the absent of the mandatory educational policy, the identification strategy could then potentially be violated in which case the strategy would not identify the effect of interest. Thus, some additional specifications that control for the interaction of years and cohort effects are explored (see section A1 of the Appendix for details).
4. Data and Sample Description

The main data used in this empirical analysis is drawn from different waves of the “Encuesta Permanente de Hogares” (EPH), an annual survey of households collected yearly in Paraguay. This survey contains information on households as well as on each of its members in the areas of labor market, education, health, and economic activity. For this study I consider the information related to education and labor market. The period of data used in this study ranges from 2002 to 2012.

The construction of the sample is based on the identification strategy proposed previously, which relies on the comparison of the outcomes of interest across different cohorts. I start by restricting the sample to individuals of age 22 across surveys, since they are more likely to have reached their desired educational attainment. The dependent variable of my model is then defined as years of formal schooling acquired by individuals at the age of 22. The sample consist of cohorts of individuals that were likely affected by the policy reform (treated group) and individuals that belong to cohorts that were not exposed to the new policy (control group). In order to appropriately define each of these cohorts, I begin from considering that the reform was first implemented in 1994, and assuming that the age of entry is six. Then, I make the following restrictions: first, the control cohorts are restricted to those individuals who were born in or before August 1985. Secondly, the fully treated cohorts are composed by those individuals that were born starting from September of 1986 and onwards. Note that the partially treated cohort is omitted from the main analysis. Later, I dismiss the individuals who failed to provide information on years of education. The final sample consists of 2,506 individuals. The control group is composed by 1,884 individuals, while the treated group consists of 622 individuals.

---

15 The raw data was obtained from the national institution of statistics, “Direccion General de Estadisticas, Encuestas, y Censos”.
16 Similar studies have highlighted that it is more appropriately to restrict the sample to young workers for the same reasons exposed here (Angrist and Lavy, 1997). Additionally, this approach helps to avoid pre-reform trends in earlier cohorts (Spohr, 2003).
17 While age of entry is 6, children are expected to turn 7 during the same academic year.
18 The expectation over the individuals that did not provide this information is that they could have less schooling, resulting in sample selection bias. This will be further discussed.
19 The final sample will be further reduced in the analysis of wages, since several individuals do not report it and are reluctant to answer to this question. This again could derive in sample selection bias.
Table 2 provides descriptive statistics for the full sample. Noticeable, Guaraní speakers have the lowest average years of formal schooling, while Spanish and bilingual speakers average years of schooling are high and very close. The gap between Guaraní and Spanish speakers is quite large. A similar pattern is seen while considering average years of schooling for wage earners. This evidences the presence of the language-based educational gap in Paraguay. Several other differences deserve attention. Household size, measured as the number of family members living in the house, an important predictor of schooling decisions, is higher for Guaraní speakers than for their counterparts. Notably, Guaraní speakers households tend to have a higher number of members relative to Spanish and bilinguals households.

Furthermore, in the case of parental education, which is represented by a variable defined as a dummy variable for parents with primary education or less, in both cases, father's and mother's education tend to be remarkably lower for Guaraní speakers. Spanish speakers tend to have more educated parents, followed by bilingual speakers. In addition to these characteristics, most Guaraní speakers are concentrated in rural areas; while a reverse situation is given for Spanish and bilingual speakers. All of this are evidencing the persistence of the intergenerational transmission of education and poverty.

What refers to labor market characteristics, the average potential experience, which is measured by subtracting years of schooling and the assumed school age of entry 6 from age, is higher for Guaraní speakers relative to Spanish and bilingual speakers\(^ 20 \). While comparing the logarithm of monthly wages (in 2002 prices), a similar pattern with average years of schooling is observed. Guaraní speakers have the lowest average monthly wages.

Another interesting feature depicted by Table 2 is related to attendance to public school. Contrary to Spanish and bilingual speakers, who attended mostly private and subsidized private schools, a high percentage of Guaraní speakers attended public schools. In developing countries, quality of education is in general very low (Glewwe and Kremer, 2006), and it tends to be even

\(^{20}\) This in line with the findings reported in other studies. For instance, Patrinos et al. (1994) observed that Guaraní speakers in Paraguay have more labor market experience than their counterparts.
lower in public schools\textsuperscript{21}. Therefore, this variable could potentially be an important determinant of educational attainment\textsuperscript{22}.

In addition, Table 3 shows means of years of education of control and treatment groups by language to reflect the underlying idea of the identification strategy. Panel A of this table compares control and treatment groups of Guaraní and Spanish speakers. From this panel, it is possible to observe an increase in average years of education and log wages for both Guaraní and Spanish speakers. The difference on rows from columns (1) and (2) illustrates the so-called language-based educational gap for cohorts before and after the implementation of the mandatory educational policy. It denotes a slight improvement in this gap.

Column (3) denotes the average gains obtained by comparing treatment and control groups for Guaraní speakers, which is subtracted from the average gains obtained for Spanish speakers. This is the Difference-in-Differences estimator, and indicates that Guaraní speakers increased the average years of schooling 0.21 years relative to Spanish speakers with the implementation of the mandatory educational policy.

5. Empirical Results

A. Effects on Educational Attainment

To evaluate the impact of the mandatory schooling policy on educational attainment, the dependent variable of the model, education, is defined as the formal years of schooling acquired by the individual at the age of 22. The reasoning of this choice relies on the assumption that 22 is a natural age at which an individual will have acquired his/her optimal schooling level\textsuperscript{23}.

Table 4, columns (1) and (2) show the results of the estimation of equation (1) for two subsamples. Column (1) shows the results for the whole sample, while column (2) restricts the sample to wage earners. In both cases, we can see that the DID estimate that compares Guaraní speakers to Spanish speakers is positive and in the order of 0.24 for the whole sample, and of

\textsuperscript{21} For instance, an usual feature in developing countries is the high absenteeism of teachers. In public schools, this tend to be frequently the case. In Paraguay, teachers' unions organize several demonstrations per year which lead to the loss of several days of classes that are not recovered.

\textsuperscript{22} I did not include this variable in the main analyses due to the its high percentage of missing or not reported information. I did consider it in the robustness check analysis, which will be discussed later on in the study.

\textsuperscript{23} Furthermore, any of the individuals in the sample go beyond college education.
0.72 years for wage earners. These coefficients denote that the average gains in years of schooling of Guaraní speakers is higher in 0.24 (0.72) years relative to Spanish speakers. However, neither of these coefficients is statistically significant at conventional levels. A similar pattern can be seen for bilingual speakers. The suggested effect of the mandatory schooling policy on bilingual speakers is higher than Spanish speakers in the order of 0.05 years for the whole sample and in the order of 0.11 years for wage earners. Nevertheless, the new mandatory schooling policy positively shifted years of schooling of all individuals regardless of the language spoken, thus ruling out the hypothesis that it has been more beneficial to the more disadvantaged language groups.

In general, these results indicate an increase of educational attainment associated with the implementation of this new mandatory educational policy. The implications of these results on the language-based educational gap are important. Considering that all individuals were benefited without distinction of language spoken, as was unexpected, no improvements were found on this gap. These findings suggest that the language-based educational gap persists and so does the need to reduce this gap.

Two factors deserve attention to explain these results. First, it is important to mention the role of language of instruction on educational attainment. Native languages, those spoken at home, which differ from the language of instruction, could potentially be a deterring factor of schooling progression in language heterogeneous countries. In this line of thought, Angrist and Lavy (1997) suggest that the persistence of instructing in a foreign language might represent a barrier, in particular to poor and rural students. In Paraguay, Spanish has been for long the language of instruction. Therefore, the findings of this study could suggest that language of instruction might still represent a barrier for Guaraní speakers.

Second, low teachers’ quality has been identified as an important factor determining poor educational outcomes in Paraguay (see World Bank 2010). With the majority of low qualified teacher concentrated in rural areas, where Guaraní speakers are main residents. Therefore, it is likely that in rural areas teachers’ proficiency in Spanish are not adequate to instruct children in that language or even to comprehend schooling materials. These effects are compounded with a lack of adequate schooling supplies.
Consequently, in the particular case revised in this study, while Guaraní speakers have benefited in a similar manner than their counterparts, the need to reduce the educational gap still continues. Then, in order to increase educational attainment of Guaraní speakers, by means to reduce the large gap existing with Spanish and bilingual speakers, more focalized actions are needed. In terms of policy implications, these results denote that an increase in mandatory schooling may not be the most effective way to enhance schooling of language-disadvantaged students. Educational policies in language heterogeneity countries should effectively target language-disadvantaged groups to obtain the desired outcomes. While more research should be undertaken to generalize these implications, this study provides an advancement on the topic.

B. Robustness Check

A difficulty that arises in this type of study is the possibility that the effects of the policy could be confounded with other macro aggregate effects or cohort effects. For instance, the implementation of this mandatory schooling policy took place immediately after the beginning of a period of democratization. It is then possible to argue that it will be difficult to separate the effects of the policy from that of the democratization. In an effort to mitigate the potential bias that could arise due to these effects, different specifications for the model, accounting for time effects, time trends, and cohort effects were explored.

B.1 Partially Treated Cohorts

As was mentioned earlier, one cohort was partially affected by the new mandatory policy. In 1998, the cohort born between September of 1985 and August of 1986 was included under the new system. This cohort was at that time in 6th grade (see Table 1 in the Appendix). Hence, for students that finished high school this would imply that they were affected by the new policy for seven years, versus the twelve years that students in any of the treated cohorts will be affected upon completion of high school. This exception allows me to use this partially treated cohort as an alternative treatment and control group. Thus, it allows me to check my DID strategy. I then compare the partially treated cohort to the treatment group, and the partially treated cohort to the

---

24 In this sense, Chiswick et al. (2000) denote that with democratization, principles and institutions are more firmly established, which leads the development process to be more spread to the broader population, rather than centering in to selected individuals or sectors.
control cohort. The idea behind this approach is that to validate the DID strategy, similar results should be found.

By comparing partially treated and treated cohorts, the results indicates that the policy reform indeed had a positive effect for all individuals of different linguistic background, which support the results obtained in the main analysis.

**B.2 Year and Cohort Effects**

The key assumption in applying the Difference-in-Differences technique is the assumption that pre-reform trends in the outcome variable would remain the same in the absence of the new policy. While it allows to identify the effects of the policy, it is a very strong assumption. In particular, if any of the confounding effects of the impact of a policy intervention takes place, then, the results could potentially be biased. For this reason, an alternative specification that includes cohorts and year effects is explored, in order to control for underlying trends. For details of the model see section A.1 in the Appendix.

By accounting for years and cohorts effects, the estimates obtained do not reveal an impact of cohort and year effects on educational attainment. Therefore, they were not included in the main analysis.

**B.3 Public School, Quantity and Quality of Education**

Several researches have indicated that often, in developing countries, quality of education is in general very low and it tends to be even lower in public schools (Glewwe and Kremer, 2006). In Paraguay this seems to hold (IDB 2004; World Bank 2010). The descriptive statistics evidence some interesting features related to public school attendance by language. While Spanish and bilingual speakers attend primarily private and subsidized private schools, Guaraní speakers attend mostly public schools.

Therefore, while the focus on this paper is mainly on quantity of education, incorporating public school in the analysis can serve two purposes. First, to examine whether the identification strategy for identifying the effects of the mandatory schooling policy on educational outcomes is validated by this approach. The mandatory schooling policy, if correctly enforced, should have higher effects on those who attended public school since they were benefited for an extra three
years of free education. Second, while an analysis of the quality of education is out of the scope of this study, the incorporation of public school attendance in the analysis could provide some insights on quality of education.

To begin this analysis, public school attendance is included in the set of regressors, as well as an interaction term between public school and being exposed to the new policy. This approach yield interesting results\textsuperscript{25}, which are reported in Table 6 for different subsamples. Notably, columns (1) and (2) indicate an increase of years of education of individuals that attended public school in the range of 0.81 to 0.87 years, which are associated with the implementation of the mandatory schooling policy. Therefore, suggesting that this policy was effective in public schools. Furthermore, these results can also be indicating that the higher years of schooling were due to an improvement in public schools. While isolating the effect of quality of schooling on educational attainment is beyond this analysis, these results provide some evidence on it. However, these results do not hold for wage earners.

Further incorporating language in the analysis, the results indicate not differential effects for Guaraní and bilingual speakers that attended public school relative to Spanish speakers. These results hold for the whole sample as well as the sample of wage earners, and validate the Difference-in-Differences results.

In summary, the analysis that incorporates public school yield positive results for those exposed by the new mandatory schooling policy that attended public school, supplementing the main analysis. These effects are not different by language groups, as suggested by previous results by applying DID.

6. Final Remarks

This study investigates the effects of a mandatory schooling policy on educational attainment and wages of individual with different linguistic background in Paraguay. By using data drawn from different waves of the "Encuesta Permanente de Hogares" in Paraguay, the study seek to

\textsuperscript{25} Results on columns (1) and (3) are obtained from the following model:

\[ S_{it} = \beta_0 + \beta_1 G_i + \beta_2 B + \delta d_i + \delta_i (d_i PS_i) + \gamma X_{ui} G_i + \gamma_2 X_{ui} B_i + \epsilon_{it} \]

while results reported in columns (2) and (4) derived from:

\[ S_{it} = \beta_0 + \beta_1 G_i + \beta_2 B + \delta d_i + \delta_i (d_i G_i) + \delta_i (d_i B) + \delta_i (d_i PS_i) + \delta_i (d_i PS_i G_i) + \delta_i (d_i PS_i B_i) + \gamma_i X_{ui} G_i + \gamma_2 X_{ui} B_i + \epsilon_{it} \]
examine whether this type of policy is effective to enhance educational attainment of the language disadvantaged students.

The results obtained from applying Difference-in-Differences econometric technique suggest an increase in years of education for all subgroups of language associated with the implementation of this policy. However, while Guaraní speakers have benefited, they were not majorly benefited as was expected. The educational gap still continues and so is the need to reduce this gap. Some explanations for these results are explored. The first one refers to language of instruction. In Paraguay, Spanish has been for long the language of instruction. For Guaraní speakers, this might still represent a barrier to schooling progression, primarily in higher educational levels. Furthermore, it is likely that in rural areas teachers’ proficiency in Spanish are not adequate to instruct children in that language or even to comprehend schooling materials. Bilingual speakers, however, do not experience these difficulties.

Finally, these findings suggest that an increase in mandatory schooling years may not be the most effective way to enhance schooling of language-disadvantaged students. In terms of policy making, these results imply that educational policies in language heterogeneity countries should effectively target language-disadvantaged groups to obtain the desired outcomes.
References


### Appendix

**Table 1 - Timeline**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5th Control</td>
<td>Sept. 80 - Aug. 81</td>
<td>7th</td>
<td>8th</td>
<td>9th</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
<td>College</td>
<td>……..</td>
<td>Labor market</td>
<td>Labor market</td>
</tr>
<tr>
<td>4th Control</td>
<td>Sept. 81 - Aug. 82</td>
<td>6th</td>
<td>7th</td>
<td>8th</td>
<td>9th</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
<td>……..</td>
<td>Labor market</td>
<td>Labor market</td>
</tr>
<tr>
<td>3rd Control</td>
<td>Sept. 82 - Aug. 83</td>
<td>5th</td>
<td>6th</td>
<td>7th</td>
<td>8th</td>
<td>9th</td>
<td>1st</td>
<td>2nd</td>
<td>……..</td>
<td>Labor market</td>
<td>Labor market</td>
</tr>
<tr>
<td>2nd Control</td>
<td>Sept. 83 - Aug. 84</td>
<td>4th</td>
<td>5th</td>
<td>6th</td>
<td>7th</td>
<td>8th</td>
<td>9th</td>
<td>1st</td>
<td>……..</td>
<td>Labor market</td>
<td>Labor market</td>
</tr>
<tr>
<td>1st Control</td>
<td>Sept. 84 - Aug. 85</td>
<td>3rd</td>
<td>4th</td>
<td>5th</td>
<td>6th</td>
<td>7th</td>
<td>8th</td>
<td>9th</td>
<td>……..</td>
<td>Labor market</td>
<td>Labor market</td>
</tr>
<tr>
<td>Partially Treated</td>
<td>Sept. 85 - Aug. 86</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td>5th</td>
<td>6th</td>
<td>7th</td>
<td>8th</td>
<td>……..</td>
<td>Labor market</td>
<td>Labor market</td>
</tr>
<tr>
<td>1st Treated</td>
<td>Sept. 86 - Aug. 87</td>
<td>1rst</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td>5th</td>
<td>6th</td>
<td>7th</td>
<td>……..</td>
<td>College</td>
<td>College</td>
</tr>
<tr>
<td>2nd Treated</td>
<td>Sept. 87 - Aug. 88</td>
<td>1rst</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td>5th</td>
<td>6th</td>
<td>……..</td>
<td>College</td>
<td>College</td>
<td></td>
</tr>
<tr>
<td>3rd Treated</td>
<td>Sept. 88 - Aug. 89</td>
<td>1rst</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td>5th</td>
<td>……..</td>
<td>College</td>
<td>College</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th Treated</td>
<td>Sept. 89 - Aug. 90</td>
<td>1rst</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td>……..</td>
<td>College</td>
<td>College</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Treated</td>
<td>Sept. 90 - Aug. 91</td>
<td>1rst</td>
<td>2nd</td>
<td>3rd</td>
<td>……..</td>
<td>College</td>
<td>College</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2 - Descriptive Statistics by Language Group

<table>
<thead>
<tr>
<th>Means</th>
<th>Guaraní</th>
<th>Spanish</th>
<th>Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schooling Years *</td>
<td>8.02</td>
<td>11.74</td>
<td>11.02</td>
</tr>
<tr>
<td></td>
<td>(3.71)</td>
<td>(3.07)</td>
<td>(3.27)</td>
</tr>
<tr>
<td>Schooling Years (wage earners)*</td>
<td>8.06</td>
<td>11.78</td>
<td>10.88</td>
</tr>
<tr>
<td></td>
<td>(3.58)</td>
<td>(2.98)</td>
<td>(3.35)</td>
</tr>
<tr>
<td>Log monthly wage**</td>
<td>12.80</td>
<td>13.55</td>
<td>13.35</td>
</tr>
<tr>
<td></td>
<td>(1.12)</td>
<td>(0.99)</td>
<td>(1.04)</td>
</tr>
<tr>
<td>Experience (years)</td>
<td>7.98</td>
<td>4.27</td>
<td>14.90</td>
</tr>
<tr>
<td></td>
<td>(3.71)</td>
<td>(3.05)</td>
<td>(3.64)</td>
</tr>
<tr>
<td>Household Size</td>
<td>6.30</td>
<td>5.53</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>(3.14)</td>
<td>(2.54)</td>
<td>(3.26)</td>
</tr>
<tr>
<td>Mother's education</td>
<td>95.28</td>
<td>54.43</td>
<td>75.48</td>
</tr>
<tr>
<td><em>(Primary or less in percentage)</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father's education</td>
<td>93.44</td>
<td>55.13</td>
<td>75.04</td>
</tr>
<tr>
<td><em>(Primary or less in percentage)</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (percentage)</td>
<td>38.16</td>
<td>57.22</td>
<td>45.96</td>
</tr>
<tr>
<td>Urban (percentage)</td>
<td>34.32</td>
<td>88.35</td>
<td>76.80</td>
</tr>
<tr>
<td>Married (percentage)</td>
<td>8.96</td>
<td>8.70</td>
<td>10.28</td>
</tr>
<tr>
<td>Public School**</td>
<td>54.7</td>
<td>41.9</td>
<td>35.5</td>
</tr>
<tr>
<td>No. of Observations</td>
<td>1250</td>
<td>575</td>
<td>681</td>
</tr>
</tbody>
</table>

Note: Statistics are underweighted. Standard errors are reported in parenthesis. The potential experience variable is approximate by age - years of schooling - 6. * Years of schooling at age 22. ** Logarithm of monthly wages in domestic currency (in 2002 prices). ***Sample reduces due to missing information.
Table 3 - Means of Education and Log(Wages) by Control and Treatment Groups and Language

<table>
<thead>
<tr>
<th></th>
<th>Years of Schooling</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (1)</td>
<td>Treatment (2)</td>
<td>Difference (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>Panel A. Guaraní vs. Spanish speakers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guaraní</td>
<td>7.64 (3.47)</td>
<td>9.71 (3.56)</td>
<td>2.08 (0.09)</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>11.20 (3.14)</td>
<td>13.06 (2.07)</td>
<td>1.86 (1.07)</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>-3.56 (0.33)</td>
<td>-3.35 (1.49)</td>
<td>0.21 (0.98)</td>
<td></td>
</tr>
<tr>
<td>Panel B. Bilingual vs. Spanish speakers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilingual</td>
<td>10.21 (3.37)</td>
<td>12.16 (2.94)</td>
<td>1.95 (0.43)</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>11.20 (3.14)</td>
<td>13.06 (2.07)</td>
<td>1.86 (1.07)</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>-0.98 (0.23)</td>
<td>-0.90 (0.87)</td>
<td>0.09 (0.64)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Statistics are underweighted. Standard errors are reported in parenthesis. * Years of schooling at age 22. ** Logarithm of monthly wages in domestic currency (in 2002 prices). The sample is restricted to wage earners.
Table 4 - Effect of the Mandatory Schooling Policy on Educational Attainment

<table>
<thead>
<tr>
<th></th>
<th>Years of Schooling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Treatment Dummy</td>
<td>0.735** (0.23)</td>
</tr>
<tr>
<td>Treatment * Guaraní</td>
<td>0.236 (0.34)</td>
</tr>
<tr>
<td>Treatment * Bilingual</td>
<td>0.050 (0.33)</td>
</tr>
</tbody>
</table>

Control variables:

- Household characteristics: yes, yes
- Parental Education: yes, yes
- Occupational Categories: no, no

$R^2$                | 0.905             | 0.916             |

No. Observations      | 2506              | 1254              |

Notes: the dependent variable is years of schooling at age 22 in columns (1) and (2). Absolute values of robust standard errors are reported in parentheses. Column (1) refers to whole sample, while column (2) refers to wage earners only. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
<table>
<thead>
<tr>
<th>Dependent Variable: Years of schooling</th>
<th>Whole Sample</th>
<th>Wage Earners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Treatment Dummy</td>
<td>0.151</td>
<td>0.161</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Treatment*Public School</td>
<td>0.812**</td>
<td>0.877**</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Treatment<em>Public School</em>Guarani</td>
<td></td>
<td>-0.415</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.58)</td>
</tr>
<tr>
<td>Treatment<em>Public School</em>Bilingual</td>
<td></td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.53)</td>
</tr>
</tbody>
</table>

Notes: the dependent variable is years of schooling at age 22 in columns (1) and (2). Absolute values of robust standard errors are reported in parentheses. All regressions are controlled for industries. * p < 0.1, ** p < 0.05, *** p < 0.01.
A1. Alternative approach: Accounting for year and cohort effects

A1.1 Educational Attainment

For the analysis of the educational attainment, accounting for possible time, age, and cohort effects, the educational outcome variable can be modeled as:

\[
S_{i\text{cl}} = \gamma_1 d_{i\text{t}} + \gamma_2 d_{i\text{c}} + \beta X_{i\text{cl}} + \sum_a \lambda_{a} d_{i\text{a}} + \sum_c \lambda_{c} d_{i\text{c}} + \sum_t \lambda_{t} d_{i\text{t}} + \varepsilon_{i\text{cl}}
\]

where, \(S_{it}\) represents the number of years of education acquired by individual \(i\) that speaks language \(l\). Let \(d_{it}\) represent the dummy variable that indicates treatment status of the individual, or more specifically, indicates whether the individual belongs to a cohort that was affected or not by the reform in the educational sector. The inclusion of a homogenous intercept is meant to capture the average outcome across individuals of the with the same language background. Then this equation indicates that in the absence of the treatment, variations in the outcome of interest across individuals is determined by a vector of demographic covariates \(X_{it}\) that are time-varying, age effects represented by \(\lambda_a\), cohort effects that are captured by \(\lambda_c\), and time effects, which are represented by \(\lambda_t\). The latter is included with the intention of controlling for underlying time trends, such as fluctuations in the macro aggregates. Finally, \(\varepsilon_{it}\) represents the error term of the equation. Furthermore, a key assumption in this econometric technique is that in the absence of this policy change, the pre-trends in the outcome variable would be the same.

From this specification it is important to note that it is not possible to identify age, cohort and survey year effects at the same time. Several actions are taken to overcome this problem. First, the outcome variable, years of education, is defined by the formal years of schooling acquired by the individual at the age of 22. Secondly, dummy variables that are generated and included in the model. Finally, homogeneous intercepts for the control and treated cohorts are included, such as the dummies will represent marginal effects to these intercepts. Then, final model is represented by equation A.2:

\[
S\text{(age 22)}_{i\text{cl}} = \gamma_1 d_{i\text{treated}} + \gamma_2 d_{i\text{control}} + \beta X_{i\text{cl}} + \sum \lambda_{c} d_{i\text{c}} + \varepsilon_{i\text{c}} \quad \forall l
\]