Are Language-disadvantaged Students at a Higher Risk of Dropping Out of School? Evidence from Paraguay

Lyliana Gayoso de Ervina\*

a Center for Economics and Human Development, Asunción, Paraguay

\*Corresponding author. Email: [lgayoso@cedeh.org.py](mailto:lgayoso@cedeh.org.py), [lyligayoso@gmail.com](mailto:lyligayoso@gmail.com)

Are Language-disadvantaged Students at a Higher Risk of Dropping Out of School? Evidence from Paraguay

Dropping out of school can have serious consequences in a child's future opportunities, widening educational attainment gaps and perpetuating inequalities. In developing countries, the mismatch between language of instruction and students’ mother tongue has been recognize as a cause for school dropout. However, research examining the drop out behaviour of students with different linguistic backgrounds is scarce. In this study, I examine if language-disadvantaged students, those who do not speak the language of instruction, are at a higher risk of dropping out of primary school in Paraguay –a bilingual, developing country. Applying survival techniques to household survey data, I estimated survival functions of each language group. The results indicate that language-disadvantaged students, Guaraní speakers, are more likely to leave primary school at any given grade after the second grade, compared to Spanish and bilingual speaking students. Furthermore, the probability of dropping out of school for Guaraní speakers reaches its peak at sixth grade. The results of the hazard models reinforce these findings, indicating that after controlling for a rich set of covariates, Guaraní speakers are more likely to leave primary school at any given year. These findings prompt further consideration of differential retention intervention approaches in countries with language diversity.

Keywords: dropout, language heterogeneity, educational gaps, survival analysis, Paraguay

JEL codes: I20, I21, I24

# Introduction

Despite the great achievements in terms of education access, high dropout rates continue to be a severe problem in the world. Dropping out of school can have serious consequences for children later functioning and adaptation and long-term opportunities, imposing high costs to society (Cairns, Cairns, & Neckerman, 1989; DesJardins, Ahlburg, & McCall, 1999; Plank, DeLuca, & Estacion, 2008; Rumberger, 1987). Moreover, school dropout can widen educational attainment gaps and perpetuate inequalities, since the non-completion of school results in children who have not acquired the most basic skills to fully function in society (Sabates, Akyeampong, Westbrook, & Hunt, 2011). To timely adopt interventions that will help students stay in school and make progress through the educational ladder, it is important to understand which students are more vulnerable to drop out and the factors that lead these students to dropping out of school.

A large branch of the literature have examined how gender, socioeconomic status, race, ethnicity, and curriculum affect dropouts in high school and middle school (S. V Cameron & Heckman, 2001; Carpenter & Ramirez, 2007; Duncan, 1994; Karl, Entwisle, & Horsey, 1997; Plank et al., 2008; Rumberger, 1983). While these studies have mostly focused on developed countries, in developing countries, factors such as gender, income and wealth, teacher and school quality, and school supply are some that appear to have a major impact in drop outs (Sabates et al., 2011). In addition, the mismatch between language of instruction and students’ mother tongue have also been recognized as a cause for school dropout (Hunt, 2008; Sabates et al., 2011).[[1]](#footnote-1) Despite this recognition, little is known about the dropout behaviour of individuals with different linguistic backgrounds under the same education regime.

This paper contributes to the literature by examining if language-disadvantaged students, defined as those who do not speak the language of instruction, are at a higher risk of dropping out of primary school compared to their language-advantaged counterparts. In addition, this study assesses whether the factors affecting language-disadvantaged dropouts are the same or differ from those affecting language-advantaged dropouts. I examine this issue in Paraguay, a country known for its high levels of bilingualism. Paraguay, as a bilingual country, has two co-official languages: Spanish and the indigenous language of Guaraní. Spanish has been the dominant language in formal sectors as well as the main language of instruction since colonial times and its mainly spoken in urban areas (Corvalan, 1984; Gynan, 2001). Guaraní is mostly spoken in rural areas and its use does not imply indigenous status, as in other countries such as Bolivia (Mortimer, 2006).[[2]](#footnote-2) Educational attainment gaps and achievement learning gaps due to language differences have been documented in Paraguay, with Guaraní speakers experiencing the lowest levels of educational attainment and achievement (Corvalan, 1984; Patrinos, Velez, & Psacharopoulos, 1994; Psacharopoulos, Velez, & Patrinos, 1994). More recently, although it is compulsory and free to attend school for nine years, as of 2017, average educational attainment of Guaraní speakers was 8.9 years, on average, compared to 12.1 and 11.1 average years of education obtained by Spanish and bilingual speakers.[[3]](#footnote-3) These statistics reflect a persistent educational attainment gap between Guaraní and Spanish speakers, which could partly be due to higher dropouts of Guaraní speaking students. However, to my knowledge, there are not studies that have investigated about the drop out behaviour of these students.

Instructing children in students’ mother tongue has been suggested as the best medium of instruction since early 1950s. Despite this, there has been a great debate over what is best in language policy in the following decades (Patrinos et al., 1994; UNESCO, 1953, 2008). Many advocate that instruction in students’ mother tongue permits progression through school levels and school completion (Corvalan, 1984). It has also been argued that it is easier for students to acquire literacy skills in their mother tongue than can later be transferred to other languages (Patrinos et al., 1994). However, others claim than teaching in students’ mother tongue instead of in the language associated with an advantaged status can actually place students at a social and economic disadvantage (Patrinos et al., 1994; Sandoval, 2012). Building on the first line of thought, I hypothesize that language-disadvantaged students face a higher risk of dropping out of school compared to their language-advantaged counterparts. Language-disadvantaged students may face language barriers at each stage of the educational ladder, which can be a deterrent to schooling progression and create early literacy problems. These problems will eventually grow, as the curriculum of higher grades require the comprehension of more complex concepts, ultimately leading students to drop out of school.

To empirically evaluate this topic, I apply Survival Analysis techniques to data drawn from the 2012 Income and Expenditure Survey of Paraguay (DGEEC, 2012). This methodology allows to estimate the probabilities of dropping out, and remedies analytical problems that arise when using standard statistical techniques. In particular, it allows to account for the temporal dimension of student departure from primary school, and censored data. The results obtained indicate that language-disadvantaged students, Guaraní speakers, face a higher risk of dropping out of primary school. Based on the results of the hazard models, Guaraní speakers are more likely to drop out of primary school even after controlling for migration and wealth, as well as other important covariates suggested by previous research. Therefore, it is likely that other unobserved characteristics of language-disadvantaged students are leading them to drop out of school, one potentially being language barriers. The analysis of the factors affecting dropouts by language group suggest that for Guaraní speakers, migrating from rural to urban areas is associated with an increased risk for dropping out, while education of a female is associated with a reduction in the risk for dropping gout for all language groups.

The contributions of this paper are important for its policy implications. First, identifying the linguistic group of students facing the higher risks for dropping out of primary school will help to identify the most vulnerable. Second, the analysis of the hazards of dropping out by grade, will inform about at which grade the risk of dropping out starts to increase and peaks. Finally, the examination of the factors affecting dropouts by linguistic groups will enable policy makers to have a better understanding of the needs of each group. These findings will help in the design and implementation of timely intervention strategies tailored to the needs of the different students’ language groups that can effectively reduce students’ dropouts and at the same time reduce inequalities in the educational sector.

# Related Literature

Empirical research in developing countries indicate that among the numerous factors influencing the decision to leave school, some belong to individuals and their households while others are external to them. Therefore, dropout predictors can be categorized into demand and supply factors (Sabates et al., 2011). Moreover, given the complexity of the dropout phenomenon, more than one single event can result in a child leaving school, and very often different factors play an interrelated role. In this section I briefly discuss selected empirical studies of dropout in developing countries.

On the demand side, household income and financial circumstances are often considered major determinants of access to education and persistence (Hunt, 2008). Access and attendance to school are associated with direct and indirect costs, such as school fees, the cost of school materials, transportation, and the opportunity cost of sending a child to school. Consequently, studies highlight the linkage between household income and dropout. Numerous studies have examined the relationship between poverty status and dropping out of school, finding that children from poor households are more likely to drop out, if ever enrolled (Brown & Park, 2002; Colclough, Rose, & Tembon, 2000). Furthermore, the likelihood of dropping out increases with time, given the increase of the opportunity cost as the child grows older (Hunt, 2008).

Income shocks can potentially trigger dropouts. Vulnerable households very often cope with income shocks by withdrawing children from school, by means of reducing costs or increasing sources of income. Nonetheless, coping strategies of poor households differ in different contexts. For instance, households participating in welfare programs, such as a conditional cash transfers, are less likely to withdraw their children from school; similarly, households with more assets are less likely to immediately pull their children out of school than households with low assets. Also, rural and poor households are more likely to be vulnerable (Hunt, 2008).

Several studies have also indicated the important role that family and households’ context play in access to education and the risk of drop out (Alexander, Entwisle, & Kabbani, 2001). Among others, education of family members, household size, and age of household members have been found to be associated with school retention and dropout. For instance, Cairns et al. (1989) examine the determinants of early school dropout finding that early parenthood is one important factor influencing the decision to leave school. Moreover, highly educated parents are more likely to value school progression and school achievement, thus their children are less likely to drop out. Similarly, higher education of the household head is associated with lower dropout rates, and more so if the head is female (Al-Samarrai & Peasgood, 1998).

A number of studies have examined the association between nutrition and health of the children, and access to education and drop out. In this sense, Glewwe & Kremer (2006) indicate that poor health status of a child can limit school participation. Additionally, it affects children academic performance, increasing therefore the probability of dropping out. For instance, Miguel & Kremer (2004) evaluate a school-based mass treatment with deworming drugs in Kenya, finding that this treatment increased participation in schools through improved health of the child. Furthermore, this treatment not only affected the treated children but also those not treated due to the reduction of disease transmission. Research also demonstrates that children suffering from chronic malnutrition, or lacking of sufficient micronutrients and energy intake are more likely to miss classes, to repeat grades, and to academically underperform, thus increasing their probabilities to drop out (Pollitt & others, 1990).

Lastly, migration can also have consequences for school attendance and dropouts. While the direction of the relationship is not certain, as it could increase or decrease school attendance and educational attainment, the literature suggests that migration from rural to urban areas frequently leads to drop out. Migration can also play an interrelated role with other factors, such as gender. For instance, Colclough et al. (2000) note that in Africa, girls are often withdrawn from school to migrate to urban areas to work as housemaids.

On the supply side, previous studies highlight that lower demand for education are exacerbated by the often insufficient lack of schools and poor school quality in developing countries (Colclough et al., 2000; Glewwe & Kremer, 2006). These supply-driven factors can affect school attendance, participation, and completion through different avenues. For example, the reduced supply of schools can increase indirect costs of education, such as transportation. In this regard, Glewwe & Jacoby (1994) examine the impact of distance to school on grade attainment and late enrolment in Ghana, finding that high travel time to school negatively affects grade attainment. Moreover, these supply-driven factors can differ by area, affecting therefore schooling decisions. In many countries, the number of schools in rural areas are much lower than in urban areas, while dropout rates are higher in rural areas.

Finally, the mismatch between language of instruction and children's mother tongue has been acknowledged as a supply side cause of drop out (Hunt, 2008; Sabates et al., 2011). The process of learning in a language that differs from the child's mother tongue in the early years of school can be exclusionary, which according to Glewwe & Kremer (2006) , is an usual characteristic of several developing countries. In an early study, Patrinos et al. (1994) states that "language selectivity exits at all levels of education and income" in Paraguay. The authors provide evidence on the effects of the role of language on educational achievement. More recently, Glewwe, Kremer, & Moulin (2009) examine the impact of the provision of free textbooks by the Kenyan government on students' academic performance. Interestingly, while these authors did not find evidence of a positive effect, their findings suggest that children were having difficulties reading the textbooks that were written in English, most students' third language. While several researchers have highlighted the linkage between language and schooling decisions, such as dropout, this relationship is still not well understood. The rest of the article is devoted to a better understanding of the dropout behaviour of individuals with different linguistic backgrounds in Paraguay.

# Methodology

Survival Analysis is a statistical method generally used to analyse data where the outcome variable is the time until the occurrence of an event of interest. More specifically, this method is known for modelling the length of time spent in a given state before transitioning to another state (Cameron & Trivedi, 2005).

Survival analysis offers multiple advantages over traditional approaches. First, traditional estimation techniques such as Ordinary Least Square (OLS) assume that residuals are normally distributed. However, time to event is rarely normally distributed, and very often not even symmetrically distributed. Consequently, assuming that time data follows a normal distribution is unreasonable (Collett, 2000). A better approach would be to assume an alternative distributional assumption for the residuals, which is the core of survival analysis.

Second, data is often censored in practice. This means that generally, subjects are not observed long enough or until time of “failure". Usually, data are analysed in a point of time when some individuals are still attending school, thus they are not observed until they drop out of school. This feature of time data renders traditional approaches inappropriate, as these cannot effectively incorporate censoring of observations, thus favouring the use of survival analysis.

Finally, traditional methods ignore time; however, variation in risk over time is the primary focus of the survival method. Considering that student progression is a dynamic process, survival analysis provides the tools to account for censored data, for risk variation over time, while at the same time permitting the use of a variety of distributions. For these reasons, survival analysis is appropriate for examining dropping out decisions.

## Survivor and Hazard Functions

Duration in a state is a nonnegative random variable, usually denoted by *T*, which has a probability density function *f(t)*. Let *F(t)* be the cumulative distribution function (cdf) of *T*, i.e., the probability that the duration is less than *t*,

. (1)

The Survivor Function , is simply the complement of the cdf shown above,

. (2)

This function reports the probability of surviving beyond time *t*. In addition to these two concepts, the hazard function is the instantaneous probability of leaving a state conditional on survival to time *t*,

, (3)

Where  . The hazard at time *t* has a direct correspondent to the risk of event occurrence at time *t*, so that for instance, when the hazard is zero there is not risk at all. The survival function, the probability density function, and the hazard function are equivalent ways of describing a probability distribution for the time of event occurrence.[[4]](#footnote-4) Finally, a related function is the cumulative hazard function, which is useful when one is interested in measuring the total amount of risk accumulated up to time *t* and can be expressed as:

, (4)

This function can be more precisely estimated than the hazard function and has an inverse relationship with the probability of survival as seen in the equation above. Given any one of these four functions, it is possible to recover the other three.

## The Kaplan-Meier Survival Estimator

The nonparametric estimator developed by Kaplan & Meier (1958) is an estimate of the survivor function *S(t)* that takes into account censored observations. Given this characteristic, its use is suggested for preliminary examination of survival data (Lee & Wang, 2003). In the context of this study, this method allows me to estimate the proportion of students who have survived up to time *t*, or equivalently, the probability of dropping out after *t*, where *t* represents years of schooling or the corresponding grade in primary education.

Suppose there are *k* distinct failure times, *t1*,………,*tk*. Failure is defined in this context as the grade or corresponding year at drop out of primary school. The Kaplan-Meier estimate at any time *t* is defined as follows:

, (5)

Where  is the number of students at the risk of dropping out at time , and is the number of students that dropped out at time . Hence, the Kaplan-Meier estimator is obtained by the product of a sequence of conditional survival probability estimators at any point in time.

## The Nelson-Aalen Estimator

The cumulative hazard function as shown in equation (4) can be estimated through the Kaplan-Meier Survival Estimator, given the theoretical relationship between  and . However, there is another non-parametric method for estimating this function, which is usually preferred given its small-sample properties. This is the Nelson-Aalen Estimator. This estimator can be defined as follows:

, (6)

where the sum is over all years of schooling less than or equal to *t*.

## The Cox-Proportional Hazard Model

The Cox-Proportional Hazard Model is a semi-parametric survival model with the advantage that it does not require the knowledge of the underlying distribution. This model is the most popular hazard model due to its computational feasibility (Cleves, Gould, Gutierrez, & Marchenko, 2010). The general form of this model can be defined as follows:

, (7)

Where  is the underlying hazard function. This function characterizes how the risk of dropping out changes with time and is also known as the baseline hazard function. In other words, this function describes the risk of dropping out for an individual with zero covariates. The other part of the hazard function, , is the relative risk and describes the effects of the covariates.

The hazard function shown in equation (7) can be further represented as,

, (7’)

Where  represents the coefficients of the covariates. The parameter estimates of the model are obtained by maximum likelihood.

The model states that the hazard ratio of any two students *j* and *m* with different covariates *xj* and *xm* is given by,

, (8)

The assumption of this model is that the hazards are proportional to each other and do not change over time, hence the hazard ratio is constant. This ratio indicates how much more likely a student is to drop out than another at any particular point in time.

# Empirical Framework

The discrete outcome of interest, dropping out of school, is assumed to be generated from an unobserved continuous variable, the household expected utility. More specifically, consumption and schooling decisions are assumed to be made by parents so as to maximize the household's aggregated expected utility. Each household decides about supporting children to continue in school by comparing future benefits and costs of attending school. If the net benefits are positive, parents support children to continue attending school, while if the net benefits are negative, children will exit school.

Considering this decision process, the dependent variable in the empirical model is the duration until the time to the relevant event. Time to event is therefore defined as schooling years attained. The event of interest, dropping out of primary school, is constructed based on information available from the survey: 1) whether the student has dropped out of school and has not returned at the time of the survey,[[5]](#footnote-5) and 2) whether the child's highest years of education attained are less than nine years of schooling. Recall that in Paraguay, primary education is compulsory and free and lasts nine years.

Then, the empirical hazard model can be written as follows,

, (9)

where *X* is a vector of regressors, *βx* is a vector of parameters to be estimated, and *µs* represents state of residence fixed effects, which measure the total impact of community and school characteristics, both observed and unobserved, on the hazard of dropping out. In this analysis the observed risk period begins when an individual starts primary school and ends when an individual experience a dropout event.

A branch of the literature on school dropout has suggested that departure from school is the end point of a long-term process in which the child has been involved. Therefore, to understand this process is necessary to consider particular personal characteristics, and the social organization of the school, home, and neighbourhood (Plank et al., 2008). Nonetheless, more recent studies have highlighted that in addition to early risk factors, dropping out of school can result from more immediate circumstances, such as health problems, or other opportunities that competes with schooling, such as the need to work. Furthermore, one can observe that within these two types of dropout approaches some factors affecting dropout belong to the individuals and their households, while others are out of their control, such as school quality.

A better understanding of dropout behaviour, in particular of individuals with different linguistic backgrounds, requires the examination of the factors that fit these two approaches. In this study, we consider a set of covariates that have been suggested in the literature as being important factors affecting the decision to drop out. The final set of covariates are summarized in the following table:

Table 1 - List of Variable Definitions and Labels

|  |  |  |
| --- | --- | --- |
| **Variable** | | **Definition** |
| *Children Characteristics* | | |
|  | age | age in years |
|  | female | 1 if female |
|  | Guaraní | 1 if speaks Guarani |
|  | bilingual | 1 if speaks Bilingual |
|  | Spanish | omitted category |
|  | firstborn | 1 if child is the older sibling |
|  | health\_ins | 1 if the child has health insurance |
|  | dropout | 1 if the child has dropout of school |
| *Family and Household Characteristics* | | |
|  | fem\_scho | years of Schooling of highest educated female in the household |
|  | sib\_scho | years of Schooling of highest educated older sibling |
|  | scholagechild | number of school-age children in the household besides the child |
|  | hh\_cct | 1 if household receives aid from a social program |
|  | hh\_remit | 1 if household receives remittances |
|  | wealth\_index | wealth index |
|  | income\_shocks | 1 if the household suffered an income shock in the past 12 months |
|  | health\_shocks | 1 if the household suffered a health shock in the past 12 months |
| *Location and Migration* | | |
|  | rural | 1 if residing in rural area |
|  | ruur | 1 if migrated from rural to urban |
|  | urru | 1 if migrated from urban to rural |
|  | state | state of residence |

The independent variables considered in the models are grouped in three categories: child characteristics; family and household characteristics; and location and migration. Child background variables include language spoken, gender, birth order, and access to health insurance. To capture language spoken by the child, two dummy variables are considered: one capturing whether the child is a monolingual Guaraní speaker, and another capturing whether the child is bilingual in Spanish and Guaraní. The omitted category is being monolingual in Spanish.

Prior studies have suggested the influence of gender status on students’ progression in school, providing mixed results. Therefore, researchers indicate that the effects of gender on school dropout are highly dependent on cultural and geographical contexts (Colclough et al., 2000). To test whether gendered effects exist on school dropout among individuals with different linguistic backgrounds, an indicator variable for female is included in the model. Furthermore, to control for possible birth effects, I consider an indicator variable capturing whether the child is the first-born child in the household. Studies that examine the effects of birth order on schooling attainment and earnings present evidence on birth order effects favouring firstborns (Behrman & Taubman, 1986; Black, Devereux, & Salvanes, 2005; Booth & Kee, 2009). In addition, a number of studies indicate the association between health status of children and dropout behaviour. In this study, I consider the variable capturing whether a child has or not health insurance. Having health insurance can assure the good health condition of a child to some extent, with the exception of extreme cases.

Family and household characteristics include education level of household members, number of siblings, access to aid, and wealth. Education level of family and household members have been suggested as potential school dropout determinants. In the hazard models, I consider the highest level of education attained by a female adult in the household, measured in years.

Additionally, the hazard models include the number of school-aged children in the household. With the inclusion of this variable, I expect to control for possible competition for household resources. Several studies are proponents of the *Quantity-quality Trade Off Hypothesis*, stating that an additional child reduces average schooling of children within the household. Then, larger households result in lower average educational attainment. However, empirical evidence is mixed. For instance, Black et al., (2005) find negative correlation between household size and children years of schooling, however, these effects dissipate after controlling for birth order effects. In contrast, Booth & Kee (2009) find that children from larger families obtain on average less years of education than children from smaller families. These results are robust to different specifications.

Income and financial circumstances are important determinants of school dropout as well. However, since income shocks can also determine dropouts, the use of income in empirical models can biased the results. Therefore, I consider the use of wealth instead of income in the empirical models. The construction of the wealth index takes advantage of the rich set of variables capturing assets owned by households. The wealth index is estimated using the principal components analysis. This index is later scaled from 1 to 10, where 10 reflects higher wealth. Among other variables included in the model, an indicator variable capturing whether the household is recipient of a welfare program is considered.

As previously discussed, immediate circumstances, such as income shocks can determine school dropout. Then, an indicator variable for income shock is included in the set of regressors. This variable capture whether the household experience a shock affecting the household income in the last 12 months. Three situations are considered: 1) if any member of the household lost his/her job, 2) if the family business filed for bankruptcy, and 3) if the household head abandoned the family. Similarly, health shocks can precipitate school dropout. Hence, an indicator variable capturing whether any of the household members became ill or had an accident in the last 12 months is considered.

Dummy variables capturing whether the child and family moved from the area where they were living when the child was born are included. One captures whether they lived in a rural area and moved to an urban area, and other captures whether they were living in an urban area and they moved to a rural area. The omitted category is no migration. Finally, with the objective of controlling for different supply side factors such as number of schools in the community, school quality, teacher quality, etc. state of residence fixed effects is also incorporated in the empirical hazard models.

# Data Sample and Description

The data used in this study are drawn from the 2012 Income and Expenditure Survey of Paraguay. This nationally representative survey contains detailed information on individuals, family, household, and dwelling characteristics.

The sample used in this study is restricted to school-aged individuals from 7 to 18 years old that started school. This data restriction obeys to the need of retrospective and current data that can determine household/children schooling decisions. Furthermore, the sample is restricted to individuals that were born in Paraguay and that reported to speak Guaraní, Spanish or that were bilinguals in both of these languages. The final sample consist of 5,481 individuals of which 448 drop out of school. A complete set of descriptive statistics of the variables used in the survival analysis is shown in Table 2 for the full sample, as well as for the sub-samples of Guaraní, Spanish, and bilingual speakers.

Table 2 – Descriptive Statistics

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables** | |  | **Full Sample** | |  | **Guaraní Sp.** | |  | **Spanish Sp.** | |  | **Bilingual Sp.** | |
|  |  |  | Mean | Std. Er. |  | Mean | Std. Er. |  | Mean | Std. Er. |  | Mean | Std. Er. |
| *Children Characteristics* | | | | | | | | | | | | | |
|  | female |  | 0.492 | 0.500 |  | 0.458 | 0.498 |  | 0.514 | 0.500 |  | 0.525 | 0.500 |
|  | Guaraní |  | 0.434 | 0.496 |  | 1.000 | 0.000 |  | ---- | ---- |  | ---- | ---- |
|  | bilingual |  | 0.256 | 0.437 |  | ---- | ---- |  | ---- | ---- |  | 1.000 | 0.000 |
|  | Spanish |  | 0.310 | 0.462 |  | ---- | ---- |  | 1.000 | 0.000 |  | ---- | ---- |
|  | firstborn |  | 0.519 | 0.500 |  | 0.471 | 0.499 |  | 0.573 | 0.495 |  | 0.536 | 0.499 |
|  | fbfem |  | 0.252 | 0.434 |  | 0.207 | 0.405 |  | 0.292 | 0.455 |  | 0.282 | 0.450 |
|  | fbru |  | 0.201 | 0.401 |  | 0.306 | 0.461 |  | 0.072 | 0.259 |  | 0.179 | 0.384 |
|  | health\_ins |  | 0.196 | 0.397 |  | 0.050 | 0.218 |  | 0.398 | 0.490 |  | 0.198 | 0.399 |
|  | dropout |  | 0.082 | 0.274 |  | 0.140 | 0.347 |  | 0.022 | 0.146 |  | 0.056 | 0.231 |
| *Family and Household Characteristics* | | | | | | | | | | | | | |
|  | fem\_scho |  | 7.717 | 4.905 |  | 5.311 | 3.743 |  | 10.979 | 4.595 |  | 8.546 | 4.472 |
|  | scholagechild |  | 1.614 | 1.423 |  | 1.895 | 1.484 |  | 1.265 | 1.258 |  | 1.560 | 1.405 |
|  | hh\_cct |  | 0.686 | 0.464 |  | 0.750 | 0.433 |  | 0.605 | 0.489 |  | 0.675 | 0.468 |
|  | wealth\_index |  | 5.516 | 2.039 |  | 4.360 | 1.713 |  | 6.979 | 1.620 |  | 5.704 | 1.789 |
|  | income\_shocks |  | 0.073 | 0.260 |  | 0.052 | 0.222 |  | 0.103 | 0.304 |  | 0.071 | 0.256 |
|  | health\_shocks |  | 0.100 | 0.300 |  | 0.109 | 0.312 |  | 0.086 | 0.280 |  | 0.100 | 0.301 |
| *Location and Migration* | | | | | | | | | | | | | |
|  | ruur |  | 0.052 | 0.222 |  | 0.055 | 0.227 |  | 0.037 | 0.189 |  | 0.066 | 0.249 |
|  | urru |  | 0.053 | 0.224 |  | 0.054 | 0.227 |  | 0.044 | 0.206 |  | 0.062 | 0.241 |
| *No. of Observations* | |  | *5,481* | |  | *2,379* | |  | *1,698* | |  | *1,404* | |

From this table, some interesting differences arise among the subgroups of language spoken. Regarding gender, the female Guaraní speaking group accounts for about 46 percent of the sample of Guaraní speaking students, while in the case of Spanish and bilingual speakers this proportion is about 51.4 and 52.5 percent, respectively. Regarding birth orders, 47 percent of the Guaraní speakers in the sample are firstborns, while in the case of Spanish and bilingual speakers, this proportion reaches 57.3 and 53.6 percent, respectively. Moreover, 20.7 percent of these firstborns are females in the case of Guaraní speakers, while 29.2 and 28.2 percent of Spanish and bilingual speakers are female firstborns, respectively. What refers to access to health insurance, only a small proportion of Guaraní speaker students have health insurance (about 5 percent). In contrast, about 40 percent of Spanish speaker students have health insurance. These statistics also highlight that the proportion of Guaraní speaking students that drop out of primary school is greater than Spanish and bilingual speakers.

Among other characteristics, large differences between schooling years of the most educated female in the household by language group can be seen. In Guaraní speakers’ households, on average, the most educated female obtains 5.3 years of schooling while in Spanish speaker households, they have about 11 years of schooling, on average. In addition, the proportion of Guaraní speaker households that participates in a social program, such as cash conditional transfers programs is about 75 percent, which is higher than their Spanish and bilingual counterparts.

What refers to the wealth index, on average Spanish speaker households’ wealth is higher than Guaraní and bilingual speaker households. This difference might be reflecting the high inequality that has persisted in the country for a long time. Finally, a higher proportion of bilingual and Guaraní speakers migrated from rural(urban) to urban(rural) areas, in comparison to their Spanish and bilingual speakers’ counterparts.

# Results

## A Preliminary Examination of Drop Outs

Table 3 allows for a closer look at the descriptive statistics of the sub-sample of students that dropped out of primary school, while Table 4 summarizes the most important reasons for dropping out reported in the survey. These two tables combined provide a better understanding of the profile of students that have dropped out of primary school.

From this sub-sample, the proportion of females that drop out is lower than males, thus in line with what has been suggested in a branch of the literature on dropouts. Furthermore, a large proportion of those who dropped out are Guaraní speakers. In addition, the proportion of students with no access to health insurance and first-born children is significantly high in this sub-sample. The later might be denoting the high rivalry among siblings at the time of competing for resources. This is reinforced by the observation that children that dropped out come from households that on average have two more school-aged children in addition to the one that dropped out. These children also come from households were the most educated female obtain on average 4.6 years of schooling.

Table 3 - Descriptive Statistics for Dropouts

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | |  | **Mean** | **Std. Er.** |
| *Children Characteristics* | | |  |  |
|  | female |  | 0.462 | 0.499 |
|  | guarani |  | 0.741 | 0.439 |
|  | bilingual |  | 0.176 | 0.382 |
|  | spanish |  | 0.083 | 0.276 |
|  | firstborn |  | 0.717 | 0.451 |
|  | health\_ins |  | 0.022 | 0.148 |
| *Family and Household Characteristics* | | | |  |
|  | fem\_scho |  | 5.281 | 3.700 |
|  | sib\_scho |  | 1.951 | 3.287 |
|  | scholagechild |  | 1.908 | 1.571 |
|  | hh\_cct |  | 0.712 | 0.453 |
|  | hh\_remit |  | 0.063 | 0.242 |
|  | wealth\_index |  | 4.142 | 1.791 |
|  | income\_shocks |  | 0.071 | 0.258 |
|  | health\_shocks |  | 0.129 | 0.336 |
| *Location and Migration* | |  |  |  |
|  | rural |  | 0.627 | 0.484 |
|  | ruur |  | 0.114 | 0.318 |
|  | urru |  | 0.049 | 0.216 |
| *No. of Observations* | |  | *448* | |

The dropouts are also characterized for participating in a welfare program, and for having low levels of assets (denoted by the wealth index) in comparison to those that did not drop out.

Table 4 – Most Important Reasons for Dropping Out

|  |  |  |  |
| --- | --- | --- | --- |
| **Reason** | | **Frequency** | **Percent** |
| *Economic Reason* | |  |  |
|  | Financially constrained | 152 | 33.93 |
|  | Needs to work | 47 | 10.49 |
| *Subjective Reason* | |  |  |
|  | Does not want to study | 110 | 24.55 |
| *Supply-side reasons* | |  |  |
|  | School is far away | 22 | 4.91 |
| *Child and Family Reasons* | |  |  |
|  | Child is ill | 21 | 4.69 |
|  | Family reasons | 61 | 13.62 |
|  | Other | 35 | 7.81 |
|  |  |  |  |
| *No. of Observations* | | *448* | *100* |

Preliminary results from nonparametric estimation suggest that Guaraní speakers are more likely to drop out than Spanish and bilingual speakers at any given year after the second year of primary school. The estimated survival functions using the Kaplan-Meier Survival Estimators are shown in figure 1, for the different language groups. Figure 1a compares the survival functions of Guaraní and Spanish speakers. As mentioned before, from this graph the differences in the risk of dropping out after a year or grade *t* are noticeable. Similarly, figure 1b compares the survival functions of Guaraní and bilingual speakers, denoting a similar pattern. These figures reveal that the gap in the risk of dropping out (or surviving) arises after second grade of primary school and reaches its highest after sixth grade of primary school.

Figure 1 – Comparison of Survival Functions

|  |  |
| --- | --- |
| (a) Guaraní vs. Spanish speakers | (b) Guaraní vs. Bilingual speakers |
|  |  |
| (c) Bilingual vs. Spanish speakers | |
|  | |

Additionally, Figure 2 shows the cumulative hazard functions for each language group. Figure 2, panel (a) shows the cumulative hazard function for Guaraní speakers.[[6]](#footnote-6) From this figure it can be seen that the cumulative hazard appears to be increasing at an increasing rate. This further implies that the hazard itself is increasing, since the hazard is the derivative of the cumulative hazard. While similar patterns can be seen for Spanish and bilingual speakers in Figure 2, panels (b) and (c) respectively, the rates are higher in the case of Guaraní speakers.

Figure 2 - Estimated Baseline Cumulative Hazard by Language

|  |  |
| --- | --- |
| 1. Guaraní Speakers | 1. Spanish Speakers |
|  |  |
| (c) Bilingual speakers | |
|  | |

## Results of Hazard Models

The semi-parametric estimation intends to shed more light on the factors associated with dropping out of primary school. Table 6 displays the results of the Cox-Proportional Hazard Model for four different models. The coefficients reported are the hazard ratios, which if significantly greater than 1 implies that a covariate is associated with an increased risk of dropping out of primary school. On the contrary, a coefficient significantly less than 1 implies a reduction in the risk of dropping out.

When modelling the event of interest, a simple baseline model is presented, which becomes more complex as more factors are included. The initial model, Model 1, contains a vector of child characteristics that includes whether the student is a female, the language spoken by the student: Guaraní or bilingual, where the omitted category is being a Spanish speaker. The model also includes a dummy variable capturing whether the child is the first-born child, interactions of first born with female and rural, and if he/she has health insurance. The results indicate that females are less likely to drop out, which is line with other studies, however, this coefficient is not statistically significant at conventional levels. Furthermore, Guaraní speaking students have a much higher dropout risk in comparison to their Spanish speaker counterparts. This coefficient is statistically significant at any conventional level. Bilingual speakers as well have a higher dropout risk compared to Spanish speakers, however, their risk is much smaller than Guaraní speakers. Being a first-born child in this model is virtually close to one and not statistically significant, while the interaction of first born with female indicates that being a first born and female increases the risk of dropping out. Furthermore, being a first born residing in a rural area is associated with an increased risk of dropping out. Finally, having health insurance decreases students' dropout risk.

Table 5 - Cox-Hazard Models of Dropping Out Estimation Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| female | 0.860 | 0.930 | 0.943 | 0.941 |
|  | (0.16) | (0.17) | (0.18) | (0.18) |
| Guaraní | 6.029\*\*\* | 2.941\*\*\* | 2.728\*\*\* | 2.729\*\*\* |
|  | (1.38) | (0.71) | (0.65) | (0.65) |
| bilingual | 2.300\*\*\* | 1.519\* | 1.412 | 1.418 |
|  | (0.54) | (0.35) | (0.33) | (0.33) |
| First-born | 0.987 | 1.451\*\* | 1.248 | 1.258 |
|  | (0.17) | (0.26) | (0.23) | (0.23) |
| fbfem | 1.099 | 1.011 | 0.962 | 0.963 |
|  | (0.23) | (0.21) | (0.20) | (0.20) |
| fbru | 1.689\*\*\* | 1.254\* | 1.604\*\*\* | 1.602\*\*\* |
|  | (0.21) | (0.17) | (0.23) | (0.23) |
| health-ins | 0.164\*\*\* | 0.311\*\*\* | 0.327\*\*\* | 0.334\*\*\* |
|  | (0.05) | (0.10) | (0.11) | (0.11) |
| fem-scho | --- | 0.920\*\*\* | 0.921\*\*\* | 0.921\*\*\* |
|  |  | (0.02) | (0.02) | (0.02) |
| scholagechild | --- | 1.096\*\* | 1.109\*\* | 1.108\*\* |
|  |  | (0.05) | (0.05) | (0.05) |
| hh-cct | --- | 0.980 | 0.969 | 0.952 |
|  |  | (0.12) | (0.13) | (0.12) |
| wealth-index | --- | 0.772\*\*\* | 0.774\*\*\* | 0.769\*\*\* |
|  |  | (0.03) | (0.03) | (0.03) |
| income-shocks | --- | --- | --- | 1.384 |
|  |  |  |  | (0.29) |
| health-shocks | --- | --- | --- | 1.256 |
|  |  |  |  | (0.22) |
| ruur | --- | --- | 2.384\*\*\* | 2.321\*\*\* |
|  |  |  | (0.44) | (0.44) |
| urru | --- | --- | 0.933 | 0.917 |
|  |  |  | (0.21) | (0.21) |
| *No. of Observations* | 5,481 | 5,481 | 5,481 | 5,481 |
| *No. of Censored* | 5,033 | 5,033 | 5,033 | 5,033 |
| *Log-pseudolikelihood* | -3,380 | -3,304 | -3,293 | -3,291 |
| *Schoenfeld Test p-values* | 0.114 | 0.140 | 0.203 | 0.165 |

*Notes:* \*p<0.1, \*\*p<0.005, \*\*\*p<0.001. Standard errors adjusted for household clusters and robust to heteroskedasticity are shown in parentheses. All regressions include state of residence fixed effects.

The results for Models 2 and 3 are shown in the second and third column of Table 6, respectively. Model 2 adds some family characteristics, while Model 3 adds some additional variables related to migration. The results obtained for the coefficient on female show that being a female student is associated with a 7 percent decrease in the hazard of dropping out in Model 2 and with a 5.7 percent decrease in the hazard of dropping out in Model 3; however, these coefficients are not statistically significant. As before, the coefficients for Guaraní speaking students are statistically significant in both models, showing that these students have a consistently higher risk of dropping out of primary school than their counterparts. These findings indicate that observed differences in dropout rates between Guaraní and Spanish speakers cannot be attributed to differences in family and household characteristics as one could suspect, since they persist after controlling for family and household backgrounds. The coefficients on bilingual students are associated with an increased risk of dropout with respect to Spanish speaking students in Models 2 and 3, nonetheless the coefficient on Model 2 is statistically significant at 10 percent, while the coefficient on this variable on Model 3 is no longer statistically significant at any conventional level.

In contrast to the results obtained in Model 1, being a first-born child is associated with an increase of students' dropout risk. While the coefficient on first born in Model 2 is statistically significant at a 5 percent level, in Model 3 it is not. These results are in line with those obtained by (Black et al., 2005; Booth & Kee, 2009). Once again, being a first born in a rural area is associated with a higher dropout risk, while having health insurance decreases students' dropout risk by almost 70 percent. The coefficient on years of schooling of the highest educated female in the household is associated with a reduced risk of dropping out. This is expected, as previous research has suggested that higher probabilities of continuing and finishing school are partly attributable to parental education, which this variable is meant to capture. The number of school-aged children in the household also plays an important role in the risk of dropout. The coefficient on this variable indicates an increased risk of dropping out, therefore, reflecting the competition for resources within the household. Furthermore, the coefficients on the wealth index in Models 2 and 3 indicate that higher levels of wealth are associated with lower risk of dropping out of primary school. Regarding migration, the hazard of dropping out is high for those who moved from rural to urban areas in comparison to those that did not move. Looking at the coefficient on moving from an urban to a rural area, it appears that this does not affect the hazard of dropping out.

Model 4 incorporates variables capturing income and health shocks in the model. The results are very similar to the ones obtained in models 2 and 3. While either of the coefficients on income and health shocks are not statistically significant at conventional levels, they indicate that experiencing an income or health shock increases the hazard of dropping out of primary school.

Regarding the satisfaction of the proportional hazard assumptions in these models, the test based on Schoenfeld residuals was performed in order to validate this assumption. The main idea underlying this test is that if the hazards are truly proportional, then the residuals should be uncorrelated with the time variable or any function of it. Under the null hypothesis of this test the correlation is zero. The p-values obtained for each model are reported at the bottom of Table 6, indicating that the proportional hazard assumption is satisfied. It is worth noting that the violation of this assumption is not necessarily bad, and in such case the results approximate the true effects on the risk of dropping out and can be interpreted as rough averages (Allison, 2010). A possible solution to the violation of this assumption is to incorporate interactions of the problematic variables and the time variable into the model.

The models presented in Table 6 share the assumption that there are not substantial differences in the effects of covariates on dropout risk for each language group. Therefore, I re-estimate Model 4 for each language group separately. The results are shown in Table 7. Being a firstborn has a larger impact on decreasing the hazard of dropping out for Spanish speakers than it does for Guaraní and Bilingual speakers. In contrast, having health insurance has a larger impact on Guaraní speakers. The results indicate that access to health insurance decreases the hazard of dropping out by 90.5 percent for Guaraní speakers and by about 60 percent for Spanish speakers. However, health insurance does not appear to affect the hazard of dropping out for bilingual speakers. In the case of the number of school-aged children, an additional child in Guaraní speaking households increases the risk of dropping out by 10 percent, and by 20 percent in bilingual households, while it does not seem to have an effect on Spanish speaking households. Furthermore, receiving remittances seems to decrease the risk of dropping out for Guaraní speakers. Finally, moving from rural to urban areas is associated with an increased risk of dropping out for Guaraní speakers.

Table 6 - Cox-Hazard Models of Dropping Out Estimation Results by Language Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Guaraní speakers** | **Spanish speakers** | **Bilingual speakers** |
| female | 0.952 | 0.944 | 0.776 |
|  | (0.21) | (0.59) | (0.32) |
| First-born | 1.617\*\* | 0.454 | 1.406 |
|  | (0.37) | (0.27) | (0.49) |
| fbfem | 0.912 | 0.961 | 1.444 |
|  | (0.23) | (0.74) | (0.64) |
| fbru | 1.296 | 3.476\*\*\* | 1.038 |
|  | (0.23) | (1.61) | (0.33) |
| health-ins | 0.104\*\* | 0.376\*\* | 0.525 |
|  | (0.11) | (0.18) | (0.27) |
| fem-scho | 0.926\*\*\* | 0.932\* | 0.875\*\*\* |
|  | (0.02) | (0.04) | (0.03) |
| scholagechild | 1.097\*\* | 1.267\*\* | 0.854 |
|  | (0.05) | (0.16) | (0.12) |
| hh-cct | 0.911 | 0.842 | 1.118 |
|  | (0.13) | (0.34) | (0.32) |
| wealth-index | 0.784\*\*\* | 0.765\*\*\* | 0.760\*\* |
|  | (0.04) | (0.08) | (0.07) |
| income-shocks | 1.355 | 0.468 | 2.507\*\* |
|  | (0.37) | (0.35) | (1.00) |
| health-shocks | 1.140 | 2.201 | 1.912\* |
|  | (0.21) | (1.27) | (0.66) |
| ruur | 2.505\*\*\* | 1.205 | 1.410 |
|  | (0.62) | (0.91) | (0.57) |
| urru | 0.887 | 0.361 | 0.990 |
|  | (0.22) | (0.44) | (0.48) |
| *No. of Observations* | 2,379 | 1,698 | 1,404 |
| *No. of Censored* | 2,047 | 1,661 | 1,325 |
| *Log-pseudolikelihood* | -2,253 | -231 | -466 |
| *Schoenfeld Test p-values* | 0.232 | 0.000 | 0.856 |

*Notes:* \*p<0.1, \*\*p<0.005, \*\*\*p<0.001. Standard errors adjusted for household clusters and robust to heteroskedasticity are shown in parentheses.

Regarding the test on the proportionality assumption, the p-values obtained for each model are reported at the bottom of Table 7 These p-values indicate that the proportional hazard assumption is satisfied for Guaraní and bilingual speakers, but not for Spanish speakers. As mentioned before, a possible solution for the violation of this assumption is to incorporate interactions of the problematic variables and the time variable into the model. This was performed for the Spanish speakers’ sample, and the results remained the same.

Overall, the results presented in this chapter suggest that Guaraní speakers are more likely to drop out than their counterparts who speak Spanish or are bilinguals in Spanish and Guaraní. The hazard of dropping out is higher for Guaraní speakers after second grade, implying that gaps in educational attainment start to emerge after this grade. The hazard of dropping out is increasing at an increasing rate over primary school for all language groups, but the rate is higher for Guaraní speakers. The results of the hazard models indicate that Guaraní speakers are more likely to drop out of primary school even after controlling for migration and wealth as well as other important covariates suggested by previous research. These findings suggest that differences in the hazard of dropout are robust, and that they might be other unobserved motives that lead Guaraní speaking students to leave primary school in the early years.

# Conclusion

This study contributes to the understanding of the drop out process of individual with different linguistic backgrounds. In particular, it examines whether language-disadvantaged students in Paraguay face a higher risk of dropping out of primary school. The findings indicate that Guaraní speakers, the language-disadvantaged students, are consistently more likely to drop out of primary school than Spanish and bilingual speakers, the language-advantaged students. The disparities on the risk of dropping out arise after second grade, and increases at a higher rate for Guaraní speakers. Furthermore, the risk of dropping out for Guaraní speakers reaches its peak at sixth grade, which implies that they are less likely to progress to seventh grade. These findings are robust across different empirical models. The results of this study, therefore, prompts further consideration of differential intervention approaches in countries with large language diversity, in order to reduce dropouts, increase educational attainment, and at the same time reduce educational gaps.

Nonetheless, the study suffers from several limitations, therefore warranting future research. First, some important variables such as age at entrance in primary school and age at enrolment at each grade are not included in the models due to the unavailability of this data. Research suggest that students who are old for a specific grade are more likely to drop out of school. Second, it is likely that unobserved characteristics of Guaraní speaker students lead them to not being able to complete primary school. Language barriers could be one of them. While this study provides evidence that Guaraní speakers are more likely to drop out of primary school, even after controlling for migration, wealth and other individual, family and household characteristics, it has not established a causal relationship between the language of instruction and dropping out. Therefore, more research is needed in order to identify if it is language barriers that are leading language-disadvantaged students to drop out of school early.

# References

Al-Samarrai, S., & Peasgood, T. (1998). Educational attainments and household characteristics in Tanzania. *Economics of Education Review*, *17*(4), 395–417.

Alexander, K., Entwisle, D., & Kabbani, N. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record*, *103*(5), 760–822.

Allison, P. D. (2010). *Survival analysis using SAS: a practical guide*. Sas Institute.

Behrman, J. R., & Taubman, P. (1986). Birth order, schooling, and earnings. *Journal of Labor Economics*, S121--S145.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children’s education. *The Quarterly Journal of Economics*, 669–700.

Booth, A. L., & Kee, H. J. (2009). Birth order matters: the effect of family size and birth order on educational attainment. *Journal of Population Economics*, *22*(2), 367–397.

Brown, P. H., & Park, A. (2002). Education and poverty in rural China. *Economics of Education Review*, *21*(6), 523–541.

Cairns, R. B., Cairns, B. D., & Neckerman, H. J. (1989). Early school dropout: Configurations and determinants. *Child Development*, 1437–1452.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Cameron, S. V, & Heckman, J. J. (2001). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy*, *109*(3), 455–499.

Carpenter, D. M., & Ramirez, A. (2007). More than one gap: Dropout rate gaps between and among Black, Hispanic, and White students. *Journal of Advanced Academics*, *19*(1), 32–64.

Cleves, M., Gould, W., Gutierrez, R., & Marchenko, Y. (2010). *An introduction to survival analysis using Stata* (3rd ed.). Stata Press.

Colclough, C., Rose, P., & Tembon, M. (2000). Gender inequalities in primary schooling: The roles of poverty and adverse cultural practice. *International Journal of Educational Development*, *20*(1), 5–27.

Collett, D. (2000). *Modelling survival data in medical research* (2nd ed.). CRC Press LLC.

Corvalan, G. (1984). Education in the Mother Tongue and Educational Achievement in Paraguay. *Prospects: Quarterly Review of Education*, *14*(1), 95–106.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review*, *18*(3), 375–390.

DGEEC. (2012). Income and Expenditure and Living Standards Survey. Dirección General de Estadísticas, Encuestas y Censos. Retrieved from www.dgeec.gov.py

DGEEC. (2017). Encuesta Permanente de Hogares (EPH). Dirección General de Estadísticas, Encuestas y Censos.

Duncan, G. J. (1994). Families and neighbors as sources of disadvantage in the schooling decisions of white and black adolescents. *American Journal of Education*, 20–53.

Glewwe, P., & Jacoby, H. (1994). Student achievement and schooling choice in low-income countries: Evidence from Ghana. *Journal of Human Resources*, 843–864.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In *Handbook of the Economics of Education* (Vol. 2, pp. 945–1017). Elsevier.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, *1*(1), 112–135.

Gynan, S. N. (2001). Language planning and policy in Paraguay. *Current Issues in Language Planning*, *2*(1), 53–118.

Hunt, F. (2008). *Dropping Out from School: A Cross Country Review of the Literature. Create Pathways to Access. Research Monograph, No. 16.* ERIC.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Karl, L. A., Entwisle, D. R., & Horsey, C. S. (1997). From First Grade Forward: Early Foundations of High School Dropout. *Sociology of Education*, *70*(2), 87–107.

Lee, E. T., & Wang, J. (2003). *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.

Miguel, E., & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159–217.

Mortimer, K. (2006). Guarani Académico or Jopará? Educator Perspectives and Ideological Debate in Paraguayan Bilingual Education. *Working Papers in Educational Linguistics (WPEL)*, *21*(2), 3.

Patrinos, H. A., Velez, E., & Psacharopoulos, G. (1994). Language, education, and earnings in Asuncion, Paraguay. *The Journal of Developing Areas*, 57–68.

Plank, S. B., DeLuca, S., & Estacion, A. (2008). High school dropout and the role of career and technical education: A survival analysis of surviving high school. *Sociology of Education*, *81*(4), 345–370.

Pollitt, E., & others. (1990). *Malnutrition and Infection in the Classroom.* UNESCO, Division of Science, Technical and Environmental Education.

Psacharopoulos, G., Velez, E., & Patrinos, H. A. (1994). Education and earnings in Paraguay. *Economics of Education Review*, *13*(4), 321–327.

Rumberger, R. W. (1983). Dropping out of high school: The influence of race, sex, and family background. *American Educational Research Journal*, *20*(2), 199–220.

Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research*, *57*(2), 101–121.

Sabates, R., Akyeampong, K., Westbrook, J., & Hunt, F. (2011). School dropout: Patterns, causes, changes and policies. *Vienna: Paper Commissioned for the EFA Global Monitoring Report, The Hidden Crisis: Armed Conflict and Education*.

Sandoval, L. O. (2012). Reforma educativa y conservación social. Aspectos sociales del cambio educativo en Paraguay. *Revista Latinoamericana de Estudios Educativo*, *42*.

UNESCO. (1953). *The Use of Vernacular Language in Education*. Paris, France.

UNESCO. (2008). *Mother Tongue Matters: Local Language as a Key to Effective Learning*. Paris, France.

1. Several studies acknowledge the existence of educational attainment gaps associated with language differences in developing countries (Angrist & Lavy, 1997; Glewwe & Kremer, 2006; Patrinos, Velez, & Psacharopoulos, 1994; Psacharopoulos, Velez, & Patrinos, 1994). [↑](#footnote-ref-1)
2. In fact, just about 2% of the Paraguayan population is ethnic indigenous, 40.3% of the total population speaks only Guaraní and about 30.4% are bilingual in Spanish and Guaraní, while monolingual Spanish speakers account for 26.8% of the population (DGEEC, 2017). [↑](#footnote-ref-2)
3. Statistics for the population between 15 and 30 years old obtained from the National Household Survey 2017 (DGEEC, 2017). [↑](#footnote-ref-3)
4. Despite thinking about the Hazard in terms of probability, technically it is not a probability function since the hazard can be greater than 1 (Allison, 2010). [↑](#footnote-ref-4)
5. It is worth noting that primary school dropout does not deter students to come back to school at some future date. [↑](#footnote-ref-5)
6. The figure depicts the cumulative hazard function for a Guaraní with all covariates equal to zero. [↑](#footnote-ref-6)