Grade Repetition and Incomplete Schooling among Adolescents in Rural Kenya

Tracyann Henry[1]

Department of Economics

Union College

e-mail: henryt@union.edu

This version: September 2004

Abstract

Common characteristics of adolescents' schooling in developing countries include tuition costs at public schools, school choice availability and at times, the choice of acquiring basic education. These characteristics, in turn structure the schools' quality and the acquisition of basic education into an investment venture at the household level. Consequently, households' schooling decisions are extended to include grade repetition and completion. This paper investigates these two decisions among a selected sample of households residing in rural Kenya. The idea of endogenous school quality and grade level as the survival time variable are introduced. Findings reveal that community and family contributions to the schools are positively related to the quality of the school. Furthermore, students from wealthier households and students attending schools with comparably higher qualities have a higher survival rate in the school system. Findings also reveal that grade 6 of the primary level poses the highest risk of early withdrawal among the adolescents.

[1]I would like to thank Mark Montgomery for his invaluable comments and his assistance in making this survey data available through the New York Research Division of the Population Council. I would also like to thank Deb Dwyer, Warren Sanderson and Andrea Tyree for their helpful comments. All errors are my own.

1 Introduction

One common characteristic of adolescents' schooling in developing countries is that it is not free (and generally, not mandatory) and a choice set of schools is available. These characteristics suggest a fresh look at the issue of school choice and student outcomes. On the supply side, varying fees and tuition inevitably lead to varying qualities of the same education product. On the demand side, parents are not only making a choice about which school to send their children to (as is the case in the US), but they are also making a decision on whether to educate their children. The education costs borne by households make this decision, whether to send their children to school, an investment issue. In this environment, parents become more concerned with the likely outcome of their children – whether they will complete the necessary levels and are there positive and attainable returns to their education.[2] If parents perceive that they will lose on this investment, then they may opt not to pursue education for their children. On the other hand, if they perceive a gain to this investment, they are expected to consider the optimal number of years as well as the choice of school based on the varying qualities.[3]

Once the decision is made to invest in the child's education, the opportunity cost of the child's time becomes important as well as the structure of the household, family income and parental perceptions about education. In terms of perceptions, gender differences – specifically in African countries – between the decision maker (household head) and the adolescents play an important role in the family's investment in education. Father's education, which may be a proxy for his perception, is found to have a significant impact on the schooling of boys and girls; the mother's perception influences only the girls's schooling (Glick and Sahn, 2000; Shapiro and Tambashe, 2002).[4] Family income and structure also play a large part in who gets educated. While increases in family income is expected to increase the likelihood of an investment being made in the child's education, family structure (domestic responsibilities to siblings) may impact negatively on this investment (Glick and Sahn, 2000). Given that investment in education does lead to positive

[2] Schultz (1989) presents arguments that returns to human capital (and hence education) is quite high in low-income countries.

[3] By taking into account the optimal number of years, parents make a decision on their children's education being limited to any of the following: some primary, complete primary, some secondary, complete secondary, etc.

[4] Further evidence show that gender differences are also extended from the household to the school level, where there are some dissimilarities with the rules and attitudes towards the education of girls and boys (Mensch and Lloyd, 1998).

returns in these low-income countries, and that family perceptions and income may have negative impacts, quite a few educational reforms have been implemented to address these issues.[5] Given the decision to invest, then the choice of school and the quality of the school are relevant to the educational outcomes faced by these adolescents. Research on schooling in developing countries have focused on a variety of educational outcomes such as student performance (Case and Deaton, 1999; Glewwe and Jacoby, 1993b), and have found that the typical school characteristics such as student-teacher ratios, do affect the performance of students on standardized tests.

While there are other student outcomes relevant to this type of investment such as late enrollment, grade repetition, and early withdrawal, this study focuses mainly on the effects of school characteristics on early withdrawal among adolescents in rural Kenya. In particular, school characteristics that have seldom been considered, such as family and community contributions are introduced in this analysis. Findings indicate that these non-traditional school characteristics have significant effects on early dropout rates, and that these characteristics are determined by household influences. This finding has important implications concerning school reform in rural Kenya. If reform concerns are about minimizing early dropout rates, then policies should be aimed at improving school characteristics that are primarily influenced by households. This can be done by either targeting the household influences to improve the school characteristics or by directly targeting the school quality (characteristics). Resources targeting households may be more effective as their positive influence will be on students already enrolled as well as potential enrollers not in the school system. Resources targeting the school quality directly will influence only those currently enrolled.

This paper continues with a review of the literature on education in developing countries with a specific focus on rural Kenya and the previous studies that have investigated adolescents' schooling outcomes there. Section 3 briefly discusses the data and its limitations, and Section 4 presents both theoretical and empirical models that will be used in the study. Results and conclusions are documented in the final sections, Sections 5 and 6 respectively.

[5]King and Orazem (1999) examined education reform in El Salvador, Columbia, Pakistan and the Philippines. Mizala and Romaguera (1999) examined education reform in Chile.

2 Literature Review

As a result of the less restricted nature of schooling in developing countries in general, issues of school choice are not limited to private and public schools but extend to choice among public schools. Research have typically focused on the direct relationship between school quality and student performance on standardized tests (Glewwe and Jacoby, 1993b; Jimenez et al., 1991; Jimenez and Sawada, 1999). This emphasis stems from a need to identify the factors that directly influence student achievement. Additionally, Glewwe and Jacoby (1993b) and Jimenez et al. (1991) in their analyses have examined jointly, student achievement and school choice as functions of household characteristics and school quality.

These studies collectively point out the significant effect of school quality on student achievement (and hence student outcome in general). In a comparison of private and public schools, private schools are found to be more efficient since they seem able to provide a higher quality of education at a lower cost (Jimenez et al., 1991). Additionally, parents are attracted to high quality schools (Glewwe and Jacoby, 1993a), and they do make this distinct choice when faced with subsidized and non-subsidized private schools (Mizala and Romaguera, 1999). However, in each study school quality has been treated exogenously.[6] This paper deviates from this typical scenario and takes into account the possible endogeneity of school quality. Ehrenberg and Brewer (1994) introduces this idea of endogenous school quality in their investigation of high school drop-out rates. Their findings indicate that parents location decisions affects student density of school districts and hence student teacher-ratios that are indicative of school quality. Furthermore, teachers' characteristics are also influenced by the salaries offered and other characteristics of the community.

Similar arguments can be made when looking at school quality and the influence of households and communities in rural Kenya. Although in rural Kenya, the location of families does not restrain school choice to only schools in the community of residence (and hence their location decisions is not a determining factor), parents and communities do have a significant impact on schools in terms of the resources made available. Resources such as subsidized housing for teachers or the

[6]The exogenous treatment of school quality may be acceptable in these essays. However, in the case of rural Kenya, it is argued here that school quality is not exogenous as parents do pay fees (and schools receive contributions from the community as well) and these fees would have an impact on the school's available resources and hence its quality.

maintenance of school buildings are significant material inputs in the education function. These inputs in turn influence the quality of the attainable pool of teachers that the school chooses from as well as the physical environment in which students are educated. Identifying these inputs would steer policies and resources toward these inputs (the primary source) instead of treating the factors they influence directly.

As in other developing countries, basic education in Kenya is neither compulsory nor free. Even though most schools (about 90%) are public schools, parents are still faced with some schooling costs. Government funding is usually limited to teachers' salaries and the principle of cost-sharing is practiced among Kenyan government schools (Ajayi et al., 1997). These costs can range from paying for textbooks to subsidizing teachers' salaries and contributing to building costs. Hence, one should expect a certain amount of variation across schools equipment and facilities (classrooms, textbooks, etc.) and hence in school quality. At the end of the primary level, parents face the possibility that their child may not be able to continue because of limited school places at the next level. Promotion to the secondary level is determined via a termination/promotion exam called the Kenyan Certification of Primary Education (KCPE) exams. About 40% of the students completing primary school are able to continue their education in a secondary school because of limited places (Ajayi et al., 1997).

In mid-1996 (May through August), the Ministry of Education and the Population Council undertook a small survey of adolescents residing in Kenya. The main aim of this survey is to examine the role of school quality in a number of educational and reproductive outcomes (Ajayi et al., 1997). Researchers involved in this project wanted to obtain a full picture of the schooling environment for Kenyan adolescents, as well as an understanding of other factors that might interfere with the learning–teaching process. Special attention is paid to the education of girls and the factors that may lead to their withdrawal from the system. Three studies have emerged from this survey that address the effect of school quality on early-withdrawal and gender differences in adolescents' schooling.

Mensch and Lloyd (1998) address the issue of gender differences in schooling among adolescents in rural Kenya. Focusing on the primary level only, their findings suggest that girls are at risk in these schools. Their performance on the KCPE are comparably worse than those of boys. Furthermore,

teachers' expectations and attitudes toward girls are discouraging and there seems to be a "double standard" in the treatment of girls. Following this article, Mensch et al. (1999) address the issue of premarital sex and dropout rates. While there are gender differences in the schooling of girls, it is also important to see whether or not this filters into their sexual activities. They conclude that girls are more likely to be involved in pre-marital sex with the likelihood increasing with puberty and age. However, girls living in female-headed households are less likely to be involved in pre-marital sex. In terms of school dropout, they do not find any significant relationship between pre-marital sex and retention, and so conclude that pregnancy is apparently not a primary reason for early withdrawal from the school system — at least for girls.

Lloyd et al. (2000) further investigate early withdrawal for girls and boys, asking whether primary school quality has any effect on this outcome. Including only individual and family characteristics, they conclude that while the risk of early withdrawal increases with age, girls face greater risk than boys; among girls, mother's lack of education increases this risk. School characteristics such as school hours and fees, teachers' credentials and gender attitudes, as well as other classroom observed factors, were found to be significant influences on the hazard of withdrawal among boys only. When using the selected group of co-educational schools, the school characteristics became significant with the girls equation.

This paper follows closely to Lloyd et al. (2000) as it addresses one issue that is considered to be problematic and at the same time, adds to the single event being considered (early withdrawal). Along with the decision to withdraw early, whether the adolescent repeats at least once is addressed, and as an extension to the model, choosing between withdrawing early and repeating is examined. The framework of a discrete-time proportional hazard model is used to examine the hazard of early withdrawal.[7] However, the typical model is adjusted so that the time variable considered is grade level with grade 1 being the initial time. Given that there is no restriction on enrollment age, using age as the time variable on which the hazard is based could lead to questionable conclusions. For illustration, consider for a moment the hazard of 10 year olds withdrawing. So for analysis, one is in fact considering the pool of 10 years olds. However, this pool of 10 year olds are not homogeneous by the fact that they could be at any grade level ranging from grade 1 to grade 5 and the decision for withdrawing from school is more than likely influenced by their current grade level.[8] Therefore,

[7]Model details are presented later in section 4

[8]Grade 5 would be the highest grade possible for a 10 year old if he enrolled when he was 6 years old and did not

using grade levels as the discrete time variable is more appropriate. Also, using grade level provides information about risky grade levels and whether there should be policies (such as mid primary exams) to encourage retention at specific grades.

This study adds to the literature by addressing separately and together the two events, early withdrawal and grade repetition in a proportional hazard model framework that uses grade level as its time variable. Additions to the literature on student outcomes in developing countries is also made through the introduction of endogenous school quality into the hazard framework.

3 Data

The final sample of adolescents extracted from the mid-1996 survey is determined through a sampling strategy involving three stages (Lloyd et al., 2000). In the first stage, 3 districts (out of a total of 50 in Kenya) are chosen to represent the range of Kenyan school environments, namely Kilifi, Nakuru and Nyeri — Kilifi representing low-income districts and Nyeri representing high-income districts. In the second stage, in each district 12 clusters are chosen (both urban and rural and also representative of the educational environment within each district). In the third stage, urban clusters are removed from the original sample of 36 clusters, and specific clusters and schools are chosen from preliminary household listings. This results in a sample of 19 clusters, 7 each from Nakuru and Nyeri and the remainder from Kilifi (Lloyd et al., 2000). The study (main survey) involves a set of surveys that are connected by individual and household identification numbers as well as school codes. These surveys are fielded for: (1) a selected group of adolescents; (2) households; and (3) a representative sample of primary and secondary schools attended by the selected group of adolescents.

The sample consists of 774 adolescents, drawn from the adolescent survey, a little more than half (58%) of them in the age group 12 – 15. Under the new school structure with 8 years at the primary level, adolescents in this age group should either be currently enrolled or have completed the primary level. The sample is well distributed across the three districts, with percentages of 32%, 38% and 30% in Kilifi, Nakuru and Nyeri respectively.

repeat any grade.

Figure 1: Summary of Sample Adolescents

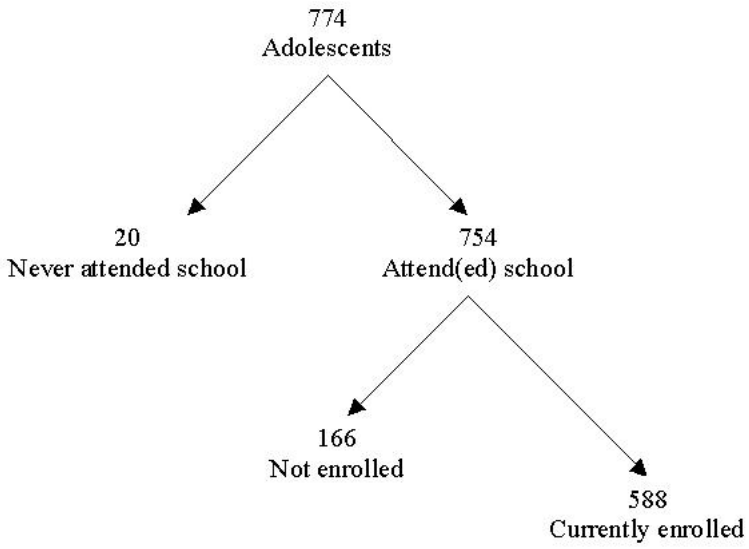


Figure 1 shows the general schooling outcome of the 774 adolescents. Once identified, adolescents are linked to the school they are currently attending or last attended (for those who are no longer enrolled). In this sample, 754 adolescents have attended or are currently attending school and 20 adolescents have never attended school. Parental inability to pay school fees is the major reason reported for never attending school — 15 out of the 20 adolescents gave this reason. The remaining 5 adolescents gave reasons that fall in the main categories of parental disapproval of school, the non-existence of economic benefits of schooling or parents believing that their adolescent is too young (or small) for school. Since the analysis focuses on the two student outcomes (behavior), early withdrawal and grade repetition, the 20 adolescents who have never attended school are excluded from the analysis.

Of the 754 adolescents, 166 are no longer attending school. Of these students, only 10.84% (18) reported having finished school as their reason for leaving school. These adolescents fall in the age group 16 to 19; 11 out of the 18 adolescents have completed at least primary level and the remaining 7 have completed the secondary level.[9] One can then conclude that the remaining 148

[9]Of the group of 11 adolescents, 9 completed primary level only while 1 each completed Form 1 and 2 at the

adolescents have withdrawn prematurely from the schooling system. These adolescents typically did not enroll late, as approximately half were enrolled in Standard 1 between the ages of 5 and 7. In regards to withdrawing early, 55% (81) gave as a reason for their early withdrawal, their inability to pay school fees. Other reasons given were lack of interest, poor performance, and being too sickly.

Now focusing on the 588 adolescents currently enrolled, it is found that their enrollment is distributed normally over a span of 4 to 12 years. Further investigation, takes into account the fact that if enrollment occurs between the ages 4 and 7 — standard 1 enrollment age is theoretically 6 years — it is probable that without repeating grades, and under the new structure, an adolescent may have completed his or her basic education between the ages 16 and 19. Focusing on the age-group 16 – 19 — among the currently enrolled adolescents, 181 (31%) fall into this age group. Among these adolescents, 122 (67%) were enrolled into Standard 1 between the ages 5 and 7, of which 30% are still enrolled at the primary level and as much as 22% are in the first year of the secondary level. Therefore, among the currently enrolled adolescents who were able to enroll on time, a sizable portion have not completed the primary and or secondary level on time. This would suggest that those adolescents who started Standard 1 at an early age and are no longer enrolled in school, are more than likely grade repeaters of at least one grade.

The effects of household wealth and consumption behavior is an important factor when examining the effects of school characteristics and school choice among this sample of Kenyan adolescents. However, household income and expenditure information is not available with this survey. To overcome this drawback, proxy variables that are typically used to measure living standards are introduced. This would in the least case, give some insight on households possible expenditure behavior. As a guide, the components of a living standard index as outlined in Montgomery et al. (2000) with the focus on household information concerning access to water, toilet facilities, housing materials and tangible assets is used. Information on the percentage of adolescents having access to general household conditions, and the assets of these households is presented in the appendix, Tables A-1 and A-2. The majority of adolescents report that public taps and rivers (or streams) are their primary source for domestic and drinking water. The houses generally have corrugated iron roofs (more than 60%) and these households typically own radios, land and livestock.

secondary level.

A total of 51 schools were surveyed, a total comprising of 36 primary schools and 15 secondary schools. The primary schools are public and the surveyed secondary schools, however, are a combination of public and private (entrepreneurial or religious) schools. With the average enrollment being quite large (approximately 200 students), it becomes relevant to look at whether or not these schools have the necessary infrastructure to support this. The main concentration is on certain attributes of the school such as the number of classrooms and the number of qualified teachers. Other features such as admission criteria for these schools are also investigated as these are possible deterrents to enrollment (or possible catalyst to late enrollment). Table A-3 in the appendix presents the fraction of sampled adolescents facing certain admission criteria as well as the fraction of adolescents attending (or attended) schools with community and family contributions.

Other school characteristics include student-teacher ratios and teacher qualifications which are presented in Table A-4 in the appendix. Classes are somewhat of the same size with an average estimated student-teacher ratio of 5. With regards to teachers' qualifications, these are grouped according to their educational achievements. The categories are defined based on their certification and are as follow: basic education, graduate education, post-secondary education, diploma, post diploma and no qualifications.[10] Overall, most of the teachers have either a basic qualification (KCPE or similar) or a post-secondary qualification (P1 - P4, S1).

Students performance on standardized tests and student-teacher ratios are arguable two of the main links between school quality and students' behavior with regards to early withdrawal and grade repetition.[11] The varied admission requirements, financial contributions and teachers' qualifications do influence the overall characteristic of the school and hence the final quality of the school.

[10]Unfortunately, information on teachers' years of experience is not available. As a result, teachers' qualifications is used as the only proxy for teachers' quality. The varied qualifications are classified as follow. Basic education include any of the following certification, CPE/KCPE, KJSE, EACE/KCE/KCSE, EAACE/KACE. Graduate education include either a graduate certification or a post graduate certification. Post secondary education include certification at any of the following levels, P1, P2, P3, P4 or S1. Diploma includes either a diploma in education or an approved teaching status. Post diploma includes either a post graduate diploma in education or a bachelor in education certification.

[11]The students' success is determined as the percentage of students obtaining a 400+ score on their exam.

4 The Model

4.1 Theoretic Approach

The theoretical model begins with the main student outcome, the likelihood of early withdrawal. This basic model is then adjusted to incorporate endogenous school quality. Grade repetition is then introduced as another student outcome and then as a possible competing risk with the decision to withdraw early.

4.1.1 Early Withdrawal

To examine the likelihood of withdrawing at the next grade level given the adolescent's current grade level, Cox's proportional hazard model and partial maximum likelihood estimation technique (Allison, 1995; Cox, 1972) is adopted. This method is adopted primarily because it gives one the freedom of not specifying a particular hazard functional form. To begin, it is assumed that the grade level variable is continuous and that no two events occur at the same time. The model is then adjusted by removing these assumptions.[12]

Consider the basic hazard model for the case of $i = 1, \cdots, n$ individuals at different grade levels, $g = 1, \cdots, G$. Assuming a linear function of the household and school characteristics, the basic hazard function for individual i is given as,

$$h_i(g) = \lambda_o(g) e^{\beta' X_i + \delta' Z_i} \tag{1}$$

where X_i is a vector of individual-specific, household and individual covariates, and Z_i is a vector of school characteristics representing school quality that each individual attending (or attended) school faces.[13] By adopting the Cox proportional hazard model, then for each occurrence of early withdrawal, the hazard that it occurred to individual i at grade level g as against occurring to any of the other individuals who are at risk in grade g is examined. That is,

[12] The nature of this analysis is such that time is not discrete and there are multiple failures at one point in time. As will be discussed later, the time variable is grade level which is discrete and it does occur that more than one individual decides to withdraw at the same grade level, hence multiple failures at the same time.

[13] As mentioned in the previous chapter, the quality variables that will be considered are students' overall performance in four subjects and student-teacher ratio.

$$L_j = \frac{h_k(g)}{\sum_{i=k}^{n} h_i(g)} \tag{2}$$

From equation 2 note that the likelihood of the j^{th} occurrence of the event happening to person k in the g grade level is equal to the hazard of this event occurring to the k person over the sum of the hazards of all individuals who are at risk in the g grade level.[14] Recalling the basic hazard function as presented in equation 1, the likelihood function is simplified further as shown below.

$$L_j = \frac{\lambda_o(g) e^{\beta' X_i + \delta' Z_i}}{\lambda_o(g) \sum_{i=k}^{n} e^{\beta' X_i + \delta' Z_i}} \tag{3}$$

This can be further simplified to,

$$L_j = \frac{e^{\beta' X_i + \delta' Z_i}}{\sum_{i=k}^{n} e^{\beta' X_i + \delta' Z_i}} \tag{4}$$

To adjust the model to account for discrete time and the occurrence of ties, Cox's modification is used Allison (1995).[15] Equation 1, the hazard equation for each individual, now becomes,

$$\frac{h_i(g)dg}{1 - h_i(g)dg} = \frac{\lambda_o(g)dg}{1 - \lambda_o(g)dg} e^{\beta' X_i + \delta' Z_i} \tag{5}$$

where $h_i(g)dg$ is a non-zero probability (Cox, 1972). The likelihood function for the j^{th} occurrence now becomes,

$$L_j = \frac{e^{\beta' S_I + \delta' W_I}}{\sum_{I=K}^{N} e^{\beta' S_I + \delta' W_I}} \tag{6}$$

Note the modifications in the likelihood function that take into account the group of adolescents withdrawing at the same time. S_I represents the sum of X_i for all individuals, I, failing at the same grade level, while W_I is the sum of Z_i over all individuals failing at the same grade level. In the denominator, the summation is done for all possible groups of individuals, K, who are at risk at the grade level g. By taking the product over all occurrences of failure, $j = 1, \cdots, m$ then the

[14] Note that to obtain this equation one must first consider a life table of the events, that starts with the first failure (withdrawal) at the earliest time (earliest grade level).

[15] Unlike age (the usual time variable), grade level is not continuous. Also, it is highly possible that at any grade level, more than one student withdraws from the school system early. The model is adjusted appropriately to take these points into consideration.

overall likelihood function is obtained. This function is in fact the partial likelihood function for the data as it take into account only those cases of failure and excludes censored cases. The log likelihood expression is given as,

$$\sum_{j=1}^{m}(\beta' S_{Ij} + \delta' W_{Ij}) - \sum_{j=1}^{m} \log[\sum_{I=K}^{N} e^{\beta' S_I + \delta' W_I}] \tag{7}$$

Maximization with respect to β for this partial likelihood function is similar to that of the ordinary likelihood function and the asymptotic properties of the estimator remains the same as it would for ordinary likelihood functions (Greene, 1997).

4.1.2 Endogenous School Quality

Now the possibility of endogenous school quality is examined. To do this, consider the specific aspects of this data, namely contributions made to the schools and the performance of the students. As noted in the previous chapter, the sampled schools receive and most likely rely on contributions from the adolescents' families as well as community contributions. Since these contributions are voluntary (not listed as admission requirements), one can assume that the families (and communities) engaging in these contributions perceive some positive value with the educational product being offered. Furthermore, families (and communities) making contributions to these schools are likely to be motivated by a desire for their children to obtain at least a basic education. It is also possible that this motivation also influences the decision for early withdrawal in that families may spend more effort in seeing that their child complete their basic education. Additionally, it has been shown in the previous chapter through simple correlations that these contributions are somewhat related to the performance of the students. As such, this unobserved motivation that influences the decision for early withdrawal (retention) also influences the performance of the students indirectly through the contributions made to the schools. It is therefore highly probable that school quality is endogenous.

In the estimation, two main variables are considered to be indicative of school quality. They are student-teacher ratio and the fraction of students successful at standardized tests (in mathematics, English, Kiswahili and agriculture). The focus is on explaining the fraction of students successful at mathematics and English.[16] The model of students' performance below assumes that the success

[16]For this study, it is assumed that the student-teacher ratio at a particular school is actually exogenous to

rates in each of the subjects are independent linear functions of community and family contributions, admission requirements and teachers' qualifications. The model, equation 8, is presented below where C is a vector of community and family contribution indicators, A is a vector of admission requirements and T is a vector of teachers' qualifications.

$$z_i = \alpha^{'} C + \phi^{'} A + \gamma^{'} T \tag{8}$$

4.1.3 Grade Repetition

The analysis extends with a similar investigation of the outcome variable, grade repetition. A similar log likelihood function as outlined in equation 7 is used to look at a single event of repeating a grade.[17] As with the decision to withdraw, maximizing this partial likelihood function with respect to β is similar to that of the ordinary likelihood function and the asymptotic properties of the estimator remains the same as it would for ordinary likelihood functions. The estimation also incorporates the possibility of endogenous school quality by including the predicted performance obtained from equation 8.

4.1.4 Competing Risks

The previous models rely on the assumption that early withdrawal and grade repetition decisions are independent of each other. Here, the possibility that grade repetition and early withdrawal are competing risks (event) in the sense that these events have different implications and the fact that either one event may occur is considered. Relaxing the previous assumption of independent events is quite applicable to this survey. In making the decision whether to repeat a grade, households making this type of education investment will likely take into account the extra expenses and time, as well as whether their child is making adequate progress. As a result, withdrawing from the school system (and possibly entering the labor market) would be among the viable options considered. To

this model and is not determined by individual and community contributions, etc. While it could be argued that contributions may attract teachers to attractive communities, the number of students per teacher is not directly influenced by these contributions. Furthermore, teachers' salaries are more likely to influence teachers towards communities, and these are in general paid by the government — which is exogeneous to this model.

[17] It would be more accurate if the estimation is modified to take into account that an adolescent may repeat more than once and hence there are multiple failures possible. However, the data is cross-sectional and hence would not be able to support multiple failures.

address this, the modifications to the previous hazard functions as pointed out in Cox and Oakes (1984) and Allison (1995) are considered. With the two competing events early withdrawal and grade repetition being possible, one can write the basic hazards model for each event is simply as follow:

$$h_{i1}(g) = \lambda_o(g)e^{\beta_1' X_i + \delta_1' Z_i}$$

and

$$h_{i2}(g) = \lambda_o(g)e^{\beta_2' X_i + \delta_2' Z_i} \tag{9}$$

where $h_{i1}(g)$ and $h_{i2}(g)$ represents the hazards for early withdrawal and grade repetition respectively. The coefficients of interest, β and δ are subscripted for each of the two events therefore suggesting that the covariates may have different effects on the different hazards. If they have the same effect on both events then the model becomes the original proportional model.

4.2 Empirical Issues

The previous theoretical early withdrawal, grade repetition and competing risks models are modified to empirical models for estimation. In the first instance, early withdrawal and grade repetition are considered independent events. As such, when examining the likelihood of withdrawing early, the event of grade repetition is censored for those who repeat. This will allow an investigation of the event, early withdrawal, among the group of adolescents at risk of withdrawing. Likewise, when examining the likelihood of grade repetition, the event of early withdrawal is censored for those individuals who withdraw early, thereby allowing the examination of the event, repeating, among the group of adolescents at risk of repeating.

The estimation investigates decisions made at the primary level, grades 1 through to grade 8. Students who graduate at grade 8 are considered censored regardless of whether they continue their schooling at the secondary level. While it would be ideal to examine the decisions throughout the basic school system — primary and secondary levels — the data is not adequate enough for this type of investigation. Recall that students enter the secondary level based on their performance on the KCPE and the availability of seats. Without additional information on individual KCPE test scores and the number of seats available at the secondary level at each point in time, it would not

be possible to identify the transition from grade 8 (primary level) to form 1 (secondary level). For example, a student withdrawing from the school system after completing grade 8 may be doing this voluntarily, may be forced to do this based on her performance on the termination / promotion examinations, or may be denied entrance due to insufficient seating at form 1.

4.2.1 Early Withdrawal

The influence of individual and household characteristics, as well as school quality on the likelihood of withdrawing at the next grade level given the adolescent's current grade level is examined. Individual characteristics include gender, age and indicators representing whether there are paternal and / or maternal siblings who are considered as part of the household. Household characteristics are determined from the 'living conditions' survey. Information on water access, toilet facilities, roofing material and assets owned do provide information on the household's standard of living and household income. Since these variables are categorial, factor analysis is used to determine an index that will capture the relevant variables. This approach to dealing with a group of categorial variables that address one issue — in this case, living standards — is typical.[18] Based on the factor loadings and the subsequent scoring coefficients on n variables(which will be discussed further in paper), household indices are created with the following general equation,

$$index = \sum_{i=1}^{n} \alpha_i \sigma_i \tag{10}$$

where α_i represent the scoring coefficient for variable i and σ_i represent the standard deviation for variable i.

Based on the scoring coefficients, one can determine whether or not the household index (or indices) will have negative or positive effects. For example, one likely expectation for the household index is that it will have a negative effect on the likelihood of withdrawing and a positive effect on the likelihood of repeating a grade. In this case, one would expect that 'wealthier' households with their child enrolled in school would be less likely to be faced with financial problems and hence

[18] Hammer and Wazeter (1993) in their examination on the effectiveness of local unions on a sample of teachers in Pennsylvania, used factor analysis to identify five aspects of union effectiveness. This was based on a survey with categorical responses, for example, choosing from 1 (strongly disagree) to 5 (strongly agree). Cooper (1983) and Bartholomew (1980) provide theoretical foundations to this approach.

less likely to withdraw their child from the school system. On the other hand, these 'wealthier' households are more likely to afford an extra year in the school system and so would be more likely to finance grade repetition.

The percentage of students successful at English and mathematics, and the student-teacher ratio are the variables included for school quality. One would expect that students exposed to higher student-teacher ratios are not receiving comparable independent attention. If they have academic problems, then these students are then more likely to withdraw early from the education system. In terms of grade repetition, the effect may not be as clear. Higher ratios on one hand, may lead to poor individual performance and an increase in the probability that the student will repeat at some point in her school career. On the other hand, if the 'class space' is quite limited (a by-product of a high student-teacher ratio), then a typical student may be discouraged from repeating and is either encouraged to withdraw or go to the next grade regardless of whether she is academically ready for this grade.

It is hypothesized here that the average student enrolled in a school with comparably higher academic success benefit from the positive peer effects (the success of other students may act as encouragement) and effective teaching. As such, one would expect that the success rates at English and mathematics to have negative effects — higher rates will decrease the likelihood of early withdrawal and / or grade repetition.

4.2.2 Endogenous School Quality

With the possibility that the success rates are a function of other community and household influences, the estimation model is adjusted so as to incorporate predicted values of these success rates. This estimation follows Heckman's two step methodology by estimating first (for each subject) the success rate as a function of household and community influences. The predicted rates are then obtained and included in the previous hazard model estimation.

Community and family contributions towards building funds and maintenance, and teachers' housing are included as explanatory variables in explaining the success rates at English and mathematics. Other explanatory variables include teachers' qualifications, admission requirements and school

management. One expects that having these non-government contributions will provide a better learning environment for the students — more or better classrooms and happier or more qualified teachers. This improved learning environment will more likely have a positive effect on the students' academic performance thereby increasing the number of students who are successful at the standardized tests. Admission requirements can in some sense act as a screening devise. Although for the most part, the student's ability or previous academic performance is not a requirement, having requirements such as fees or pre-school attendance may discourage those households that value schooling less.[19] As a result, the students that are enrolled in these schools with attendance requirements are more likely students who have a higher valuation on the schooling system and are more likely to perform better.

School management in terms of parents' involvement is expected to have some effect on the quality of the school. A higher fraction of parent involvement is representative of the value placed on the school (and the schooling system). One can imagine that these parents are more likely to want to optimize their children's performance at the school. This motivation (or desire) is expected to have positive effects on their children which could be reflected in their performance and hence the overall success rates at the school. The hypothesis here is that school management will have a positive effect on the success rates in English and mathematics.

4.2.3 Grade Repetition

The estimation model used to explain the likelihood of withdrawing early is now extended to the likelihood of grade repetition. To examine the hazard of repeating, only the first incidence of grade repetition is considered. While students do have multiple grade repetition, the data is cross-sectional, thereby limiting hazard estimations to one-time failures as against multiple failures. Furthermore, the first incidence of grade repetition is chosen as the failure event (as against the second or last incidence of grade repetition). Consider the first incidence of grade repetition. The student is allowed to repeat courses with the aim that she will improve in these subjects. Her future decisions concerning possible second and subsequent incidences of grade repetition are then in some part influenced by her first grade repetition experience. Her improvements may reduce the

[19] One can see that households that value schooling less are more likely to withdraw their child early from the schooling system.

likelihood of repeating another grade. As such, the likelihood of her second and subsequent grade repetitions occurring are influenced by the first event of grade repetition. Given that the data does not provide time-varying information that may capture these changes, using the first incidence of grade repetition as a failure is appropriate.

4.2.4 Competing Risks

The final estimation now considers grade repetition and early withdrawal as competing risks. Given that one of the main reasons for early withdrawal is financial difficulties, it is likely that when at the end of each level, if the student is not doing well enough to progress to the next grade, then the household is faced with the option of withdrawing their child from the schooling system or allowing the child to repeat. Furthermore, these events can be considered competing as the occurrence of one event precludes the occurrence of the other — which is characteristic of events being competitive Allison (1995) . If the student withdraws, then she cannot repeat the grade, and if she chooses to repeat the grade, then at that point she cannot withdraw.

To test whether the events are competing risks, a third model is estimated where the failure event is withdrawing early or repeating a grade. The test is graphical and involves examining the survival functions of the three models; models 1, 2 and 3 with early withdrawal, grade repetition, and early withdrawal or grade repetition as the respective failure events. The test will be based on their proportions and is as follow. If the models remain proportional over grade levels and are parallel, then these events (early withdrawal and grade repetition) are competing risksAllison (1995). On the other hand, if they are not parallel, then the null hypothesis that they are competing risks is rejected. This test is quite simple and does not involve the covariates. Hence, the test and conclusions are not affected by estimation issues.

5 Estimation and Results

5.1 Early Withdrawal

By estimating the the likelihood of early withdrawal, the investigation focuses on explaining the probability of withdrawing before going to the next grade level. In other words, the interest lies

in determining the grade level at which the adolescent is at risk most. Such information will improve on the efficiency and effectiveness of policies that target retention in the basic schooling system. This estimation will determine the individual and household characteristics that influences this behavior as well as identify school characteristics that also influence this behavior. These significant characteristics once identified, provides a viable tool to execute polices that are more global.

Table A-5 in the appendix presents the results from the probability of withdrawing at the next grade level. Typical household and individual characteristics are included here and an indicator as to whether or not the adolescent has information from the school survey data. The results indicate that girls and older adolescents have a relatively higher hazard and so are less likely to go on to the next grade level. Adolescents from relatively poorer households, measured by roofing material and source of domestic water, also have a lower survival rate and are more likely to withdraw early.

These results are quite preliminary as only selected household characteristics are included as proxies for the household's living standard. It is quite possible that there are omitted household characteristics that are relevant to this estimation. As mentioned in previously, one way to account for the contribution of these characteristics, household indices (factors) are estimated using factor analysis. Because of the wide range of household conditions and assets identified, appropriate proxies for household income (wealth) to be used in the analysis is determined by looking at the correlation coefficients between these measurements of living standards and the two student outcomes identified – early withdrawal and grade repetition. The correlation coefficients along with their significance levels are presented in the appendix, Table A-3. Variables as listed in Table 1 are included in the factor analysis. Eleven factors were created and the following graph, Figure 2 plots their eigenvalues. This graph helps in determining how many factors should be retained.

From this graph, note that the difference between the first and second factor is greater than one.[20] This large difference implies that only one factor (the first factor) should be retained and that the other eigenvalues are sampling noise. Therefore, the first factor is retained as the household index. The scoring coefficients are obtained for this factor and are presented below in Table 1.

This household index, along with the school characteristics, success rates at English and mathematics, and student-teacher ratios are now included in the estimation of the hazard model on early

[20]The other eigenvalues have differences that are at most 0.5.

Figure 2: Principal Factors for Household Index

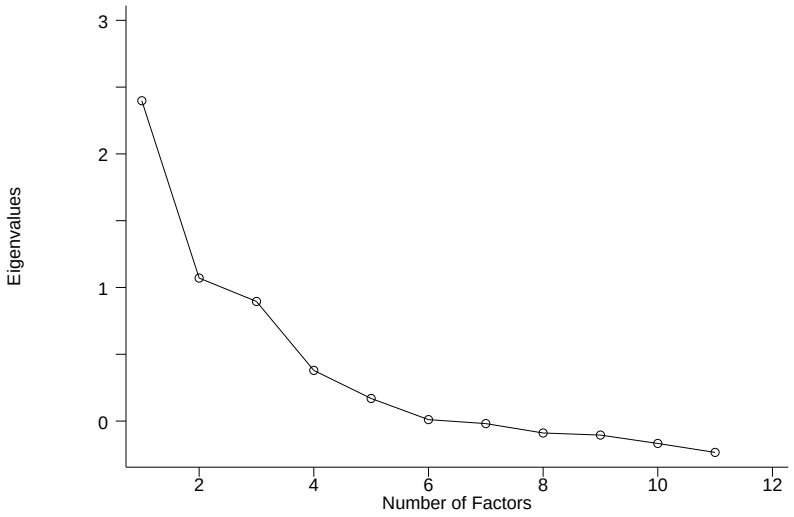


Table 1: Household Index Scoring Coefficients

Variable	**Scoring Coefficient**
Domestic water piped into the house	-0.17376
Domestic water from a well	0.02223
Domestic surface water	0.07206
Tradition pit	-0.02080
Grass roofing	-0.41182
Iron roofing	0.45705
Own a radio	0.01917
Own a bicycle	0.02435
Own land	0.03705
Own cattle	0.03739
Own a stall	0.01352

withdrawal. The results are presented in the first three columns representing Model A in Table 2 below.

Looking at the adolescents' characteristics, note that although the effect of age is not significant, older students are less likely to withdraw from the school system early.[21] While girls are found

[21]There should be some level of caution with this finding concerning the age of the children. Two are two possible

Table 2: Probability of Early Withdraw

	Model A			Model B		
	haz. ratio	coeff.	s.e.	*haz. ratio*	coeff.	s.e.
Female	1.409	0.343*	0.205	1.290	0.255	0.206
Age	0.489	-0.716	0.805	0.541	-0.615	0.805
Age squared	1.032	0.032	0.025	1.029	0.028	0.025
Household Characteristics						
Paternal Siblings	0.949	-0.052*	0.031	0.944	-0.058*	0.032
Maternal Siblings	1.008	0.008	0.006	1.007	0.007	0.006
Household Index	0.603	-0.506**	0.109	0.579	-0.547**	0.107
Time to obtain water	0.997	-0.003	0.005	0.995	-0.005	0.005
School Quality						
Success rate: English	2.889	1.061	1.498			
Success rate: mathematics	0.396	-0.927	1.586			
Student-teacher ratio	1.070	0.067*	0.038	1.091	0.087**	0.041
Predicted: English				30.186	8.013**	3.082
Predicted: mathematics				38.880	-10.158**	3.551
Missing Variables Indicators						
English, mathematics	1.419	0.350	0.647	1.655	0.504	1.052
Student-teacher ratio	1.012	0.012	0.735	0.465	-0.765	1.067
Nobs		519			519	
No. of failures		105			105	

* 10% level of significance

** 5% level of significance

to have a higher likelihood of withdrawing, it is interesting to note that the presence of paternal siblings in the household decreases the likelihood of withdrawing early. Also, note that having a

reasons for this result. First, it could be reflecting the fact that older children are more mature and probably further in the education system and hence they have a higher survival rate. Secondly, it could be the case that these children started school late and if early withdrawal is more prominent at the later grades, then one would not be able to see their intention at this time. To investigate this further, should be extended to take into account the adolescent's enrollment age.

higher household index decreases the likelihood of withdrawing early and this effect is significant at the 5% level. This means that students with higher levels of family income as reflected through the household index, are more likely to remain in the schooling system and not withdraw early. This hints at the idea that households financial status remains a strong determinant in the student's schooling even after she has enrolled in a school.

In terms of school quality, the results reveal that adolescents exposed to lower student-teacher ratios have a relatively higher survival rate. This finding is not surprising as one would imagine that lower student-teacher ratios will lead to – at least in theory – more time being spent per student. The other indicators of school quality, the fraction of students being successful in standardized English and mathematics test are not significant but their relationship with the probability of withdrawing in the next period is quite interesting. The performance of students in English has a positive effect. This implies that students attending schools with a comparatively higher overall performance in English have a higher likelihood of withdrawing at the next grade level. On the other hand, performance of students in mathematics has a negative effect, implying the opposite effect – that attending schools with a comparatively higher performance in mathematics decreases the likelihood of withdrawing at the next grade level. As mentioned in the previous section, performance in these tests are considered an indication of the school's quality. Higher success rates would imply higher quality schools. This, along with the findings of Model A, Table 2 implies that performance in mathematics may be more important than performance in English when it concerns retention in school. However, since the results are not significant, conclusions are not made at this point. Instead, the possible endogeneity of these two school indicators are explored next.

5.1.1 Endogenous School Quality

A two-stage estimation procedure very similar to that outlined by Heckman's two-stage is performed. First, predictors of the student performance variables are obtained from the regression of success rates on community and household contributions, attendance requirements and school management. The results from this estimation is presented in Table 3.

In summary, the estimation results reveal that community and family contributions do have an overall significant and enhancing effect on the performance of students. This finding is not surprising as contributions towards teachers' housing and building funds are likely to attract teachers with

Table 3: Explaining Student Performance in English and Mathematics

	English		Mathematics	
	coeff.	std. err.	coeff.	std.err.
Community Contributions				
Building funds	0.454**	0.029	0.295**	0.032
Building maintenance	-0.717**	0.040	-0.626**	0.045
Teachers' housing	0.579**	0.035	0.459**	0.039
Family Contributions				
Building funds	0.295**	0.028	0.262**	0.031
Teachers' Qualifications				
Basic Degree	0.005**	0.001	0.005**	0.001
No Qualification	0.007	0.006	0.011*	0.007
Admission Requirements				
Fees	-0.039**	0.019	-0.029	0.021
Pre-school Attendance	0.116**	0.042	0.116**	0.047
School Management				
Log (Yearly Income)	0.048**	0.013	0.034**	0.014
Fraction of parents (board)	-0.025	0.029	-0.040	0.033
Fraction of parents (pta)	0.108**	0.023	0.092**	0.026
Constant	-0.425**	0.127	-0.251*	0.142
Nobs	492		492	
Adjusted R^2	0.581		0.436	
Predicted (Y) - mean	0.205		0.247	

* 10% level of significance

** 5% level of significance

comparably higher quality and provide adequate classrooms, two qualities that one would expect to add positively to student's performance. Community contributions to building maintenance on the other hand has a negative and significant effect on students overall performance in English and mathematics. One possible explanation for this effect is that if the community is providing maintenance funds (as against receiving aid through government sources), these funds may not be adequate. If this is the case, then the likelihood of deteriorating buildings increases and this has the potential to impact negatively on students performance. However, since it is not clear from the survey whether community maintenance funds are received in lieu of government assistance, no strong conclusion will be made on this finding.

Teachers' qualifications also seem to have some effect. Having a higher fraction of teachers with basic qualifications (as against at least a teaching diploma) increases the success rate in English and mathematics standardized tests. While this finding is somewhat surprising, one possible explanation is that the qualification covariate is also reflecting years of experience and / or turnover rates. It is possible that the schools that have a higher percentage of their teachers having only the basic qualifications are also the schools that have teachers with more years of experience and also very low turnover rates. One would expect these attributes to have a positive effect on students' performance and it may be that it is this effect that dominates. Unfortunately, the data does not provide information on these attributes. In terms of admission requirements, schools that require pre-school attendance have higher success rates and schools that require fees, have lower success rates.

In terms of school management, although the effect is not significant at the 10% level, having a higher fraction of parents on the school board actually affects the students performance negatively. Note however that parents involvement in Parent Teachers Associations do have a positive and significant effect on the school's success rate in English and mathematics. From these estimations, the average predictions of student performance indicate a 21% and 25% success rate in English and mathematics respectively. The predictions are used in the second step of the two-step methodology. Estimation of the probability of early withdrawal which includes these predicted covariates is presented in Table 2, Model B. Generally, it is noted that the previous effects of adolescents and household characteristics remain the same. Adolescents residing in households with higher standard of living and with more paternal siblings are less likely to withdraw early from the school

system. Student-teacher ratio continues to be positive and significant. Students attending schools with higher student-teacher ratios are more likely to withdraw from the school system at the following grade level.

Interestingly enough, the effects of student performance — predicted success rates in English and mathematics — are now significant and the signs remain the same as they were before. Adolescents attending schools with higher fractions of students successful at mathematics have a comparatively lower hazard for early withdrawal. However, the results concerning student performance in English is puzzling. It suggests that students attending schools with higher student performance in English, have a higher hazard of withdrawing early from the school system. Since the relationship between withdrawing and performance in English has been consistent (whether it is significant or not) this finding questions whether or not performance in English is a reasonable indicator of overall school quality.

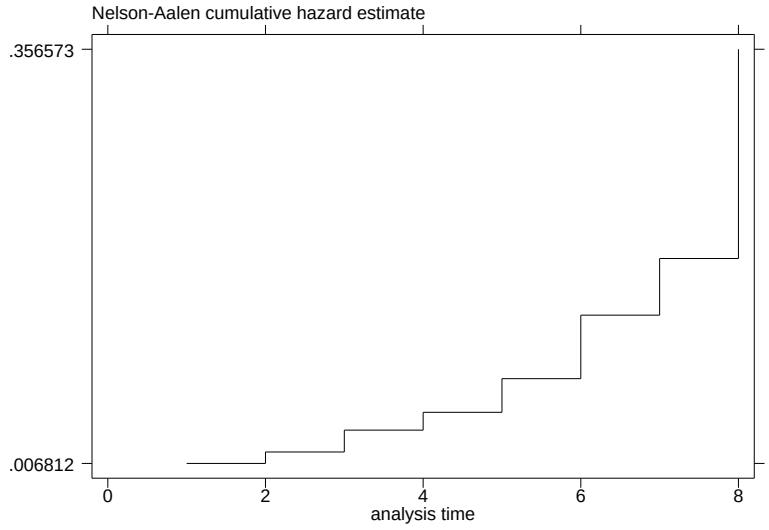
The above findings reveal that school characteristics do have an effect on the survival rate of adolescents enrolled at the primary level in rural Kenya. Student-teacher ratios continue to have a positive effect — increasing the student-teacher ratio the average student faces by 10 will increase her likelihood of withdrawing at the next level by almost 1%. Also, students enrolled in schools with a higher fraction of students being successful at mathematics have a higher survival rate in the school system. Furthermore, these success rates are partially explained by community and family contributions, hinting at the notion that community involvement — which is indicative of their valuation on schooling — is a key ingredient in the success rates at these schools. With these findings, the focus is now on investigating which grade level is at the highest risk. Table 4 below presents the estimated baseline survivor function at each grade level for the above two-step estimation.

From Table 4 the probability of surviving just beyond grade 1 is 88% and this probability decreases steadily as the adolescent progresses through the school system. Notable declines in the estimated survival probabilities occur between grades 4 and 5 (a decline of 16%) and between grades 5 and 6 (a decline of 17%). These changes in the probability of surviving hints at the possibility that grades 4 and 6 are likely the riskier grades. The Nelson-Aalen cumulative hazard function, Figure 3 provides another examination of this hypothesis as it represents the cumulative hazard estimates for the entire group of non-censored adolescents.

Table 4: Estimated Baseline Survivor Function for Early Withdrawal

Grade Level	Estimate	Difference
1	0.88312	
2	0.73800	0.14512
3	0.59571	0.14229
4	0.48838	0.10733
5	0.33197	0.15641
6	0.15783	0.17414
7	0.08154	0.07629
8	0.00295	0.07859

Figure 3: Nelson-Aalen Cumulative Hazard Function



From the graph, note that the hazard rate seems to be increasing at a relatively uniform pace between grades 1 to 4. The hazard rate increases sharply at grade 5 and at grade 6. Once again, this suggests that the riskiest grade is grade 6, or the transition from grade 6 to grade 7.

5.2 Grade Repetition

Grade repetition among the adolescents in rural Kenya is now examined. Initially, one can imagine that the two factors — household income and school effectiveness — are the main determinants in the decision to repeat a grade. Conditional on the student's ability, if the school is comparatively effective, as reflected by its quality, then the average student's likelihood of repeating a grade should be comparatively lower. Now, if the student is faced with the decision to repeat a grade level, her decision can be interpreted in more than one way. If the event is considered as an external shock to the student's pre-determined education plan, then with education being not free, one can imagine that household income becomes a factor as repeating a grade implies financing an additional year of school. As such, one may observe income — as measured by the household index — to increase the likelihood of repeating a grade at the next level. On the other hand, if the possibility of this event is actually factored into the initial decision process such households are somewhat financially prepared for such an event, then households income in this situation may not have a significant effect on the likelihood of this event.

The investigation of this even begins with the hazard model of grade repetition is estimated and presented in Table 5.

From Table 5, school characteristics, student-teacher ratio and success rates, although they have the similar effects as with early withdrawal, these effects are not significant. Household characteristics seem to dominate the likelihood of repeating a grade level. Having more maternal siblings present in the household reduces the hazard of repeating at the next grade level, and students living in households with higher standard of living are less likely to repeat at the next grade level.

The results are quite interesting once they are compared to those from the estimation of early withdrawal. Recall from the first estimation that students residing in wealthier households and exposed to schools with comparably higher quality (higher success rates and lower student teacher ratios) are at a lower risk of withdrawing early from the school system. However, with regards to the risk of repeating a grade, the quality of the school does not play a significant role (although the signs of the coefficients are as expected)and students residing in wealthier households have a lower risk of repeating a grade. Together, this suggests that students with wealthier families are at a lower risk of repeating a grade or withdrawing early. Also, focusing on the effectiveness of the school, as reflected by the school quality, there is some evidence that the quality of the school

Table 5: Probability of Grade Repetition

	haz. ratio	coeff.	std.err.
Female	0.896	-0.110	0.298
Age	6.736	1.908	1.514
Age squared	0.921	-0.082	0.052
Household Characteristics			
Paternal sibling	1.005	0.004	0.006
Maternal sibling	0.849	-0.164**	0.064
Household Index	0.574	-0.555**	0.155
Time to obtain water	0.995	-0.005	0.008
School Quality			
Student-teacher ratio	1.059	0.058	0.045
Predicted rate: English	0.019	-3.985	5.147
Predicted rate mathematics	10.155	4.607	5.999
Missing Variables Indicator			
English, mathematics	0.000	-17.750**	0.771
Nobs		519	
No. of failures		50	

* 10% level of significance

** 5% level of significance

influences the decision to repeat and the decision to withdraw early.

The question then that follows from these findings is whether these two events are competing risks. That is, when faced with the decision to withdraw early from the system, do these households see grade repetition as a viable alternative? The consistent effect of the school quality variables and household characteristics, regardless of whether they are significant, suggest that this may be the case.

5.3 Competing Risks

To examine whether grade repetition and early withdrawal are competing risks, the estimation model is adjusted to consider the first occurrence of either even as a failure. The results from this estimation are presented in Table 6 below.

Table 6: Probability of Combined Failure (grade repetition or early withdrawal

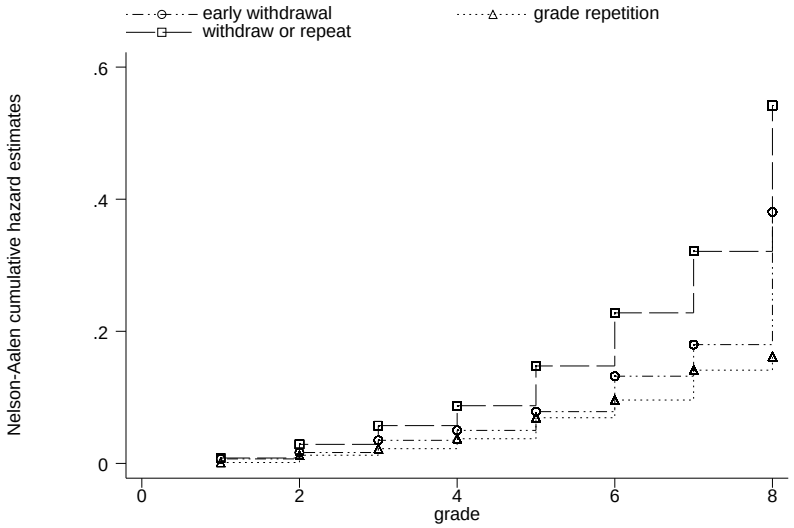
	haz. ratio	coeff.	std.err.
Female	1.187	0.172	0.166
Age	0.189	-1.664**	0.572
Age squared	1.055	0.054**	0.018
Household Characteristics			
Paternal sibling	0.997	-0.003	0.006
Maternal sibling	1.005	0.005	0.006
Household index	0.585	-0.536**	0.088
Time to obtain water	0.996	-0.004	0.004
School Quality			
Student-teacher ratio	1.085	0.082**	0.031
Predicted rate: English	25.214	3.227	2.592
Predicted rate: mathematics	0.014	-4.267	2.950
Missing Variables Indicator			
English, mathematics	1.230	0.207	1.038
Student-teacher ratio	0.942	-0.060	1.041
Nobs		519	
No. of failures		150	

* 10% level of significance

** 5% level of significance

The results show that older students are less likely to withdraw early or repeat a grade level. This effect decreases with age. Also, the findings indicate that students with wealthier households are

Figure 4: Graphical Test for Events Independence



at a lower risk of repeating a grade or withdrawing early form the school system. In terms of school quality, note that the student-teacher ratio has a positive effect on the combined events. Students exposed to schools with higher student-teacher ratios are at a higher risk or repeating a grade or withdrawing from the school system early. Although the effect of the success rates are not significant, the signs are as one would expect. In terms of mathematics, students attending schools with higher fractions of successful students are comparatively face a less risk with regards to grade repetition and early withdrawal.

To complete this section, a graphical test is carried out to determine whether these events are indeed competing risks. Figure 4 presents the graphs used to test this hypothesis — that grade repetition and early withdrawal are competing risks. The test involves comparing the Nelson-Aalen cumulative hazard functions for the estimations, namely: (1) the hazard of early withdrawal, (2) the hazard of grade repetition, and (3) the hazard of early withdrawal and grade repetition. If the two events are competing risks, then the should have similar proportional hazards. As such, the three functions should be parallel to each other (StataCorp, 2001). If this is true, then the hypothesis is accepted — the events are competing risks.

Results from the graph suggest a rejection of the hypothesis that grade repetition and early withdrawal are competing risks. While the curves seem parallel to each other at grades 1 through to grade 4, note that the proportions at the higher grade levels are not comparable. One should

however take some caution with this result. This test is quite fragile and the rejection of the null hypothesis may be based on other underlying assumptions used in developing the models.

6 Concluding Remarks

A couple of points based the overall findings emerges from this study. The results here concur with previous studies in finding that family characteristics do affect student outcomes such as early withdrawal. Students residing in comparably wealthier households (as measured by the index) are less risky with regards to early withdrawal. Additionally, it found that school characteristics do influence withdrawal decisions. Students attending higher quality schools as measured by student-teacher ratios and success rates, are at a lower risk for early withdrawal. Further investigation of the school characteristics, mainly the success rates, reveal that household and community contributions are significant determinants.

Combining these results together presents a significant and important conclusion on schooling in rural Kenya. On one hand, communities, based on the value they place on education, provide financial assistance — building funds, teachers' housing — to the schools in the communities. This community involvement has a positive influence on the quality of the schools which is reflected in the success rates of the students taking the standardized tests. Students attending these quality schools are at a lower risk of withdrawing early. Additionally, students from wealthier households are also at a lower risk of withdrawing early. Note then that the main influence on adolescents survival in the school system is the availability of funds. Students from poorer households are at risk and students attending poorer schools (in terms of community and family contributions)are also at risk.

Furthermore, it was also found that students in grade 6 are at a higher risk of withdrawing early from the school system. There could many reasons of this grade level being the turning point. It would be interesting to know the curriculum at this level and also at higher levels. It is possible that grades 7 and 8 are preparation grades for the termination / promotional examinations, KPCE and that those students who do not intend to advance to secondary school make the decision to withdraw at this point in time. However, this is just a hypothesis, and sufficient information is not available to explore this further. Overall, this finding that community and family contributions play a significant role in determining student retention is quite important as it suggest the possibility of

a vicious cycle surrounding families, communities and education. If the community in general does not put a high value on schooling (in comparison to other investments) — as measured by their contributions, then this will affect the quality of the school and individual household decisions concerning enrollment, grade repetition and retention. This will again aggregate to the community's value on the school system. Additionally, the non-mandatory clause in the education system may exacerbate this cycle.

Designing policies geared towards improving student outcomes must then be done cautiously. For instance, the provision of vouchers made available to households in the less affluent communities, may not be effective as the non-mandatory clause in the education system does influence household's education decisions. In lieu of suggesting radical policies such as creating a free and mandatory schooling system at the primary and secondary levels, a more moderate policy would be one that focuses on the students who have decided to attend school. Such polices would then be focused on retention and repetition among other outcomes. Given the results on grade repetition and early withdrawal at grade level 6, a policy that targets this grade, either by providing vouchers or scholarships to to qualifying students would likely be more effective in increasing retention rates. A possible design of the voucher or scholarship program is one that awards the voucher or scholarship based on academic performance in the previous years. This will not only encourage the students who are currently enrolled, but also, it may attract those who have opted for no schooling due to finances.

Finally, although the results were not too supportive of the idea that grade repetition and early withdrawal are competing risks, one should not rule out the idea completely. It is possible that both events are explained by different hazard functions. Note that in reality, the event of grade repetition is different as it allows multiple 'failures' while the event of early withdrawal only allows one 'failure'. If it is the case that while independent, both events require different hazard functions, then the graphical test would not be appropriate. Additionally, this analysis depends on the idea that both events are independent of each other in the sense that observed and unobserved factors influence the decisions of either event in the same manner. If this is not the case, and there are unobserved factors that may lead to the student to choose one event over the other, then the events are not independent and the analysis of competing risks is flawed. To account for this, Allison (1995) suggest the developing models that take into account this level of dependence. As future

research, it would be interesting to test whether these events are competing risks after developing the appropriate models.

References

Ajayi, A., W. Clark, A. Erulkar, K. Hyde, C. B. Lloyd, C. B. Mensch, C. Ndeti, B. Ravitch, E. Masiga, and S. Gichaga (1997). Schooling and the experience of adolescents in kenya. Nairobi, Kenya: Government of Kenya Ministry of Education and Population Council.

Allison, P. D. (1995). *Survival Analysis Using The SAS System: A Practical Guide*. Cary, NC: SAS Institute Inc.

Bartholomew, D. J. (1980). Factor Analysis for Categorical Data. *Journal of the Royal Statistical Society Series B (Methodological) 42*(3), 293–321.

Case, A. and A. Deaton (1999). School Inputs and Educational Outcomes in South Africa. *The Quarterly Journal of Economics 114*(3), 1047–1084.

Cooper, J. C. B. (1983). Factor Analysis: An Overview. *The American Statistician 37*(2), 141–147.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B(Methodological) 34*(2), 187–220.

Cox, D. R. and D. Oakes (1984). *Analysis of Survival Data*. London, England: Chapman and Hall.

Ehrenberg, R. G. and D. J. Brewer (1994). Do School and Teacher Characteristics Matter?: Evidence from High School and Beyond. *Economics of Education Review 13*(1), 1–17.

Glewwe, P. and H. Jacoby (1993a). Delayed Primary School Enrollment and Childhood Nutrition in Ghana: An Economic Analysis. World Bank Living Standards Measurement Study, Working Paper No. 98, Washington DC: The World Bank.

Glewwe, P. and H. Jacoby (1993b). Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana. *Journal of Human Resources 29*(3), 843–864.

Glick, P. and D. E. Sahn (2000). Schooling of Girls and Boys in a West African Country: The Effects of Parental Education, Income and Household Structure. *Economics of Education Review 19*(1), 63–87.

Greene, W. H. (1997). *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall Inc.

Hammer, T. H. and D. L. Wazeter (1993). Dimensons of Local Union Effectiveness. *Industrial and Labor Relations Review 46*(2), 302–319.

Jimenez, E., M. E. Lockheed, and V. Paqueo (1991). The Relative Efficiency of Private and Public Schools in Developing Countries. *The World Bank Research Observer 6*(2), 205–218.

Jimenez, E. and Y. Sawada (1999). Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program. *The World Bank Economic Review 3*, 415–441.

King, E. M. and P. F. Orazem (1999). Evaluating Education Reforms: Four Cases in Developing Countries. *The World Bank Economic Review 3*, 409–413.

Lloyd, C. B., B. S. Mensch, and W. H. Clark (2000). The Effects of Primary School Quality on School Dropout among Kenyan Girls and Boys. *Comparative Education Review 44*(2), 113–243.

Mensch, B. and C. Lloyd (1998). Gender Differences in the Schooling Experiences of Adolescents in Low-Income Countries: The Case of Kenya. *Studies in Family Planning 29*(2), 167–184.

Mensch, B. S., W. H. Clarke, C. B. Lloyd, and A. S. Erulkar (1999). Premarital Sex and School Dropout in Kenya: Can Schools Make a Difference? Working Paper No. 124, Policy Research Division, Population Council.

Mizala, A. and P. Romaguera (1999). School Performance and Choice: The Chilean Experience. *Journal of Human Resources 35*(2), 392–417.

Montgomery, M. R., M. Gragnolati, K. A. Burke, and E. Paredes (2000). Measuring Living Standards with Proxy Variables. *Demography 37*(2), 155–174.

Schultz, T. W. (1989). Investing in People: Schooling in Low Income Countries. *Economics of Education Review 8*(3), 219–223.

Shapiro, D. and O. B. Tambashe (2002). Gender, Poverty, Family Structure, and Investments in Children's Education in Kinshasa, Congo. *Economics of Education Review 20*(4), 359–375.

StataCorp (2001). *Stata Reference Manuals, Release 7.0, Volumes 1–7*. College Station, TX: Stata Corporation.

7 Appendix

Table A-1: General Household Conditions

	Total Number of Adolescents
Source of Domestic Water	
Water piped into house	81(10.74%)
Public Tap	245(32.49%)
Well with pump	26(3.45%)
Well without pump	110(14.59%)
Lake or pond	12(1.59%)
River or stream	168(22.28%)
Purchased water	5(0.66%)
Rainwater	50(6.63%)
Other	28(3.71%)
Missing information	29(3.85%)
Source of Drinking Water	
Same source as domestic	704(93.37%)
Other sources	21(2.79%)
Missing information	29(3.85%)
Toilet Facilities	
Own flush toilet	3(0.40%)
Shared flush toilet	2(0.26%)
Traditional pit	177(23.47%)
Ventilated pit	65(8.62%)
No modern facility	4(1.57%)
Missing Information	503(66.71%)

Table A-2: General Household Assets

	Total Number of Adolescents
Roofing Material	
Grass or Thatch	243(32.33%)
Corrugated Iron	476(63.13%)
Other	2(0.27%)
Missing Information	33(4.38%)
Household Appliances	
Electricity	26(3.47%)
Radio	487(64.59%)
Television	60(7.96%)
Refrigerator	7(0.93%)
Other Assets	
Bicycle	176(23.34%)
Land	563(74.67%)
Livestock†	488(64.72%)
Market Stall or Shop	77(10.21%)

†: livestock includes cattle, goats or sheep

Table A-3: General School Conditions

	N	mean	s.d.	min	max
Admission Requirement					
Community Resident	520	0.038	0.192	0	1
Age or Birth Certificate	520	0.306	0.462	0	1
Age only	520	0.242	0.429	0	1
Readiness	520	0.065	0.247	0	1
Health	520	0.031	0.173	0	1
Examination	520	0.438	0.497	0	1
Fee Payment	754	0.168	0.375	0	1
Pre-School Attendance	754	0.651	0.477	0	1
School Uniform	520	0.277	0.448	0	1
Books	520	0.117	0.322	0	1
Desk	520	0.069	0.254	0	1
Previous School Documents	520	0.006	0.076	0	1
Other	520	0.012	0.107	0	1
School Management					
Yearly Cost (KSh$)	520	1826.094	2071.935	400	20450
Fraction of parents (school board)	520	0.598	0.345	0	1
Fraction of parents (PTA)	512	0.516	0.321	0	0.95
Community Contribution					
Building Maintenance	754	0.086	0.281	0	1
Building Funds	754	0.093	0.290	0	1
Educational Supplies	520	0.090	0.287	0	1
Teachers' Housing	754	0.057	0.232	0	1
Family Contributions					
Building Maintenance	520	0.002	0.044	0	1

Table A-4: School Characteristics

	N	mean	s.d.	min	max
Percentage of Teachers by Qualification					
Basic	754	12.936	11.166	0	41
Graduate	520	0.277	1.780	0	17
Post-Secondary	520	17.590	8.569	0	40
Diploma	520	0.575	1.697	0	13
Post Diploma	520	0.242	1.648	0	16
None	754	0.430	1.034	0	8
Student-teacher Ratio	754	5.28	4.45	0	13
Percentage of Students Successful					
English	754	0.137	0.198	0	0.697
Mathematics	754	0.165	0.201	0	0.788
Kiswahili	754	0.213	0.245	0	0.867
Agriculture	754	0.165	0.184	0	0.625

Table A-5: Probability of Early Withdraw (without school characteristics)

	haz. ratio	**coeff.**	**std. err.**
Female	1.447	0.370∗	0.200
Age	1.305	0.267∗∗	0.060
School: same cluster	0.965	-0.036	0.221
School information indicator	1.645	0.498	0.493
Household Characteristics			
Water source: surface	0.611	-0.492∗	0.266
Time to obtain water	0.998	-0.002	0.005
Toilet: traditional pit	1.470	0.385	0.520
Radio	0.731	0.314	0.205
Roof: iron	0.489	-0.716∗∗	0.227
Nobs		519	
No. of failures		105	

∗ – 10% significance level

∗∗ – 5% significance level

Table A-6: Correlation Coefficients between Student Outcomes and Household Variables

	Early Withdraw		Repeat Primary	
	ρ	sig.	ρ	sig.
Water Source				
Pipe	0.0907∗	0.0138	-0.0963∗	0.0082
Well	-0.0171	0.6434	0.0527	0.1486
Surface	-0.1182∗	0.0013	0.0367	0.3144
Time to access water				
(min, max) = (1, 140)	-0.0920∗	0.0362	0.0554	0.2039
Toilet Facility				
Traditional pit	0.0782∗	0.0340	0.0106	0.7711
Roof Material				
Grass or Thatch	0.1179∗	0.0014	-0.1126∗	0.0020
Corrugated iron	-0.1350∗	0.0002	0.1261∗	0.0005
Other Indicators				
Electricity	-0.0585	0.1207	-0.0906∗	0.0148
Radio	-0.0707	0.0602	-0.0015	0.9685
Television	-0.0196	0.6035	-0.0492	0.1866
Refrigerator	-0.0132	0.7260	-0.0451	0.2258
Bicycle	0.0701	0.0625	0.0119	0.7493
Land	-0.0510	0.1757	0.0682	0.0667
Cattle	-0.0647	0.0857	0.1457∗	0.0001
Stall	-0.0104	0.7840	-0.0089	0.8107

* 5% significance level