CENTER FOR LABOR ECONOMICS
UNIVERSITY OF CALIFORNIA, BERKELEY
WORKING PAPER NO. 29

# Ethnic Diversity and School Funding in Kenya

Edward Miguel*
University of California at Berkeley
emiguel@econ.berkeley.edu

December 2000

**Abstract**

The impact of ethnic diversity on the provision of local public goods and collective action in Africa remains largely unexplored. To address this gap, this paper explores the relationship between ethnic diversity and local primary school funding in rural western Kenya. The econometric identification strategy relies on the stable, historically determined patterns of ethnic land settlement in western Kenya. The main empirical result is that higher levels of local ethnic diversity is associated with sharply lower primary school funding and worse school facilities in western Kenya. The theory examines school choice and funding decisions when pupil mobility among schools is limited by land market imperfections and ethnic divisions, the relevant case for rural Africa, and predicts that local pupil transfers may lead to upward bias in OLS estimates of the impact of ethnic diversity. This theoretical prediction is confirmed in the data.

**JEL Classification: H4, I2, O1**
Keywords: Local public goods, ethnic diversity, social capital, education, Kenya, Africa

* George Akerlof, Alberto Alesina, Abhijit Banerjee, Mary Kay Gugerty, Caroline Hoxby, Lawrence Katz, Michael Kremer, T. Paul Schultz and seminar participants have provided valuable comments. I am especially indebted to the staff of ICS Africa, Sylvie Moulin, Michael Kremer, and Elizabeth Beasley for their generosity. I gratefully acknowledge financial support from the U.S. National Science Foundation and the MacArthur Foundation. All errors are my own.

## 1 Introduction

An increasing number of studies have suggested that ethnic and other social divisions may be important determinants of economic outcomes. Recent research suggests that higher levels of ethnic diversity are related to low provision of local public goods across United States municipalities,[1] and that ethnically diverse societies may be prone to corruption, political instability and slow economic growth,[2] due to political conflict and lack of cooperation across ethnic groups. However, although the relationship between ethnic diversity and economic outcomes is likely to be particularly important for sub-Saharan Africa – which is the most ethnically diverse and the poorest region in the world, and has suffered from a series of destructive ethnic conflicts in recent years[3] - the impact of ethnic diversity on local collective action and public good provision in sub-Saharan Africa remains largely unexplored empirically.

This paper addresses these issues through an examination of the relationship between ethnic diversity and local primary school funding in western Kenya. Textbooks, classrooms, desks, and other important educational inputs are locally funded by parents in Kenya. The theoretical model of school choice and local funding developed in the paper illustrates how pupil mobility among schools may complicate the estimation of a causal relationship between ethnic diversity and school funding levels. Two assumptions distinguish it from existing school choice models, which examine U.S.-style settings.[4] First, households are not residentially mobile due to frictions in rural land markets and the costs of moving across ethnic

[1] Alesina, Baqir, and Easterly [1999] find that high levels of ethnic diversity are associated with up to 25 percent lower local funding for schools and other public goods in U.S. municipalities. Goldin and Katz [1997] argue that public secondary schooling expanded slowly in ethnically diverse U.S. school districts from 1910 to 1940.

[2] In cross-country empirical work, Mauro [1995] finds that national ethnic diversity is significantly related to poor bureaucratic performance and political instability across countries, and Easterly and Levine [1997] conclude that ethnic diversity has been a principal cause of slow economic growth in Africa during the post-colonial period.

[3] The 1994 Rwandan genocide – in which perhaps 500,000 people were massacred, often by their neighbors – is perhaps the most tragic example of ethnic conflict in Africa (Des Forges [1999]). The notion that ethnic diversity has been an impediment to development in sub-Saharan Africa is also found in the popular media: "Many of the world's problems stem from the fact that it has 5,000 ethnic groups but only 190 countries. ... The region with perhaps the most intransigent ethnic rivalries is sub-Saharan Africa, which has about 1,300 language groups in 42 countries, the boundaries of which were imposed by the colonial powers with little regard for ethnicity" (Rodger Doyle, *Scientific American*, September 1998).

[4] Refer to Benabou [1993] and Fernandez and Rogerson [1996].

boundaries, limiting school choice to local schools within walking distance of the home. Second, the school choice decision is largely driven by exogenous variation in school academic quality, which can be thought of as teacher competence, honesty, and motivation; teachers are centrally appointed to primary schools in Kenya, typically with minimal local control. The theory illustrates pupil sorting patterns among schools in this setting, and implies that pupil sorting will bias ordinary least squares estimates of the impact of ethnic diversity on school funding, as good quality schools attract ethnically diverse pupil populations from the surrounding area.

To address the potential sorting bias characterized in the model, the relationship between ethnic diversity and school funding in western Kenya is identified using local residential ethnic composition as an instrumental variable for school ethnic composition. Ethnic land claims in western Kenya were established in the 1800s and have been largely stable during the past century. The use of historically determined ethnic settlement patterns to estimate the impact of local ethnic diversity on public good provision constitutes an improvement over recent estimates of the impact of ethnic diversity from the United States: the high rate of residential mobility observed in the United States complicates the interpretation of coefficient estimates on ethnic diversity, since unobserved aspects of school quality or individuals' taste for education may be correlated with local ethnic composition.

The main conclusion of the empirical analysis is that local ethnic diversity is associated with sharply lower local school funding and worse facilities in ninety rural Kenyan primary schools. The drop in funding associated with the change from complete ethnic homogeneity to median school ethnic diversity is approximately 25 percent of average local school funding, and this relationship is robust to the inclusion of geographic, socioeconomic, demographic and teacher quality controls. The theoretical prediction of an upward bias in OLS estimates of the impact of ethnic diversity on school funding is confirmed in the data, and a variety of evidence suggests that attenuation bias due to measurement error - an alternative explanation for the bias - is not driving the pattern of OLS and IV coefficient estimates.

These findings may shed light on the relationship between ethnic diversity and the provision of other local public goods in Kenya, which are funded through similar public finance mechanisms. The results

also contribute to the current debate on the sources of poor African economic performance, since primary school quality is an important determinant of human capital accumulation in poor countries and human capital investment has been associated with subsequent economic growth.[5] The region of Kenya under study is fairly typical for Africa in terms of income and ethnic divisions, suggesting that the results may have implications for local collective action in other less developed countries. Finally, the results point to the need for further research on policies that build cooperation and trust – or "social capital" – across ethnic groups in societies with prominent social cleavages.

The rest of the paper is structured as follows: the setting in western Kenya is described in Section 2, the theory of school choice and its empirical implications are presented in Section 3, and the empirical results are discussed in Section 4. Section 5 summarizes the results and examines possible implications for economic growth and public policy.

## 2 The Setting

### 2.1 Primary Schools in Busia and Teso Districts, Kenya

Busia and Teso districts are poor agricultural regions in Kenya's Western Province. In 1996, the original Busia district was split in two: Teso district is the northern part of the original Busia district, and Busia district is the southern part. Figure 1 presents the location of Busia and Teso districts on a map of Kenya. The combined population of Busia and Teso districts in 1989 was 401,658, and their total area 1,652 square kilometers (Government of Kenya [1994]). The average daily wage for agricultural labor in this area is approximately 0.85 U.S. dollars, which is low by Kenyan standards (Gugerty and Miguel [2000]). Busia and Teso district primary schools are typical for Kenya in terms of educational attainment: the combined Busia district ranked twenty-sixth of fifty districts on 1995 national primary school exams (Glewwe, Kremer, and Moulin [1998]). The material poverty of primary schools in Busia and Teso is striking. Few classrooms for the lower grades have desks, so most pupils sit on the dirt floor; pupil

[5] Refer to Krueger and Lindahl [2000] for evidence on the cross-country relationship between education and growth.

textbooks are rare and chalk is in short supply; and classes are sometimes held outside due to a lack of permanent classroom structures.

Both the central government and local school committees play important roles in Kenyan primary school finance. The national Kenya Ministry of Education pays teacher salaries, while school committees raise funds locally for books, chalk, classrooms, and desks. Although the teacher salaries and benefits paid by the central government account for most primary school spending – over 90 percent of total primary school spending according to the author's calculations – a reduction in local funding could have an important impact on educational outcomes if local inputs and teachers are complements in educational production.[6]

The school headmaster collects most local school funds from parents in the form of annual school fees, which are set by each school's primary school committee. Local community members who do not have children in the school do not typically participate in the school committee, and they are not expected to pay school fees. The school committee is composed of twelve members directly elected by the parents of schoolchildren, and it typically meets at least three times per year to discuss school finances and plan school projects. School fees in Busia and Teso districts ranged from 200-500 Kenya shillings (5-12 U.S. dollars) per family in 1995. The second source of local primary school funding – accounting for approximately one-third of local funding in western Kenya – are village fundraisers called *harambees* in Swahili, at which parents as well as other community members meet and publicly pledge their financial support for a planned school investment project, such as the construction of a new classroom. *Harambees* are an important source of public finance throughout Kenya, accounting for 40 percent of total local

[6] In addition to its school finance role, the national Ministry of Education assigns teachers and headmasters to primary schools. Survey evidence indicates that eighty percent of teachers in Busia and Teso districts are assigned to teach in their native "home" area (Gugerty and Miguel [2000]). Although it is not unknown for school committees to lobby for the transfer of poorly performing teachers and headmasters, the Ministry of Education generally assigns teachers with minimal local participation. Masara [1996] contains a discussion of the policies of the national Teachers' Service Commission (TSC). I thank Mary Kay Gugerty, Sylvie Moulin and Robert Namunyu for their observations on these and other issues in Kenya.

expenditures on primary schools and other local public goods (Wilson [1992]). Prominent public figures – such as Members of Parliament – often attend fundraisers and make contributions (Barkan [1994]).[7]

The author conducted structured field interviews with twelve primary school headmasters in this area during June 2000 – at six schools in Busia district and six schools in Teso district, in both ethnically diverse and homogeneous areas – and their responses indicate that schools employ a variety of methods to encourage school fee and *harambee* payment in western Kenya. By far the most common method of pressuring parents to contribute is to publicly announce the names of parents who are late with payments during parent and school committee meetings: ten of the twelve headmasters reported that they regularly read out the names of parents late with payments at school meetings, which they believe embarrasses the parents and encourages other community members to pressure them to pay. The headmaster of Nanderema Primary School in Busia district stated that parents who have already "paid [their fees] become very bitter, and help us collect money from others," and the headmaster of Buduta Primary School in Teso district stated that parents who have paid "bully those who have not paid." Headmasters often employ additional forms of social pressure to encourage prompt payment of *harambee* contributions or school fees, including sending letters to the homes of parents late with fees, asking local church leaders to encourage payment during sermons, and making personal visits to the individual homes of debtors accompanied by the local Chief.

The responses also indicate that headmasters often threaten to expel pupils whose parents do not make *harambee* contributions or pay school fees. However, few pupils are ever permanently expelled for non-payment: eleven of the twelve headmasters do send children home if their parents are late with payments, but the children are typically allowed back into school in a matter of weeks or months even if parents have not yet paid the entire fee.

[7] Although data on total *harambee* collections exist for these schools, there is insufficient information on the identity of contributors to establish the proportions provided by parents, other community members, and local politicians.

## 2.2 Ethnic Groups

Busia and Teso districts are moderately ethnically diverse: the largest ethnic groups are the Luhya (67 percent of the sample), Teso (26 percent), and Luo (5 percent).[8] The Luo and Teso are Nilo-Saharan ethno-linguistic groups with pastoralist traditions, and the Luhya are a Bantu (Niger-Kordofanian) group. Luhyas are the majority ethnic group in Busia district and Tesos are numerically dominant in Teso district, although there are significant minority communities on both sides of the Busia-Teso border. The Luhya are composed of the Khayo, Marachi, Nyala, and Samia subtribes, among others.[9] It is unclear if the Luhya should be considered one or many ethnic groups since certain Luhya subtribe dialects are mutually unintelligible, there are historical rivalries between subtribes, and the notion of a single Luhya ethnic group is recent, originating in the colonial period (Government of Kenya [1986]; Were [1967]). For this reason, ethnic diversity measures that classify Luhyas as one group or as several distinct groups are both considered in the empirical analysis.

Ethnicity is perhaps the primary cleavage in Kenyan political and social life, and several violent ethnic clashes – claiming hundred of lives – have occurred in both central and coastal regions of the country during the past decade, most notably in advance of the 1992 and 1997 national presidential polls (Ndegwa [1997]). Although there has been no ethnic violence in Busia or Teso district during the post-colonial period, there is evidence that ethnic tension is common in a variety of settings and may have an adverse impact on local collective action. Government of Kenya anthropologists write that: "The Teso, as a minority group surrounded by people with whom they were on belligerent terms in the not so distant past, tend to have rather strained relations with their neighbors. … It is not uncommon for Teso and Luhya to come to blows in places of work. … These attitudes make it difficult to design development

[8] The sources of data presented in this section are discussed in section 4. School children are generally taught in their vernacular (native) language in western Kenya through grade three, although Swahili is usually the medium of instruction in ethnically diverse areas. Starting in grade four, classes are principally conducted in English.

[9] Other Luhya subtribes in Busia and Teso districts include Bukusu, Dakho, Kabras, Marama, Sukha, and Wanga. Other non-Luhya ethnic groups in this area include Kalenjin, Kikuyu, Masaai, Somali, Tachoni, and Taita.

projects for the District as a whole which would require inter-ethnic cooperation" (Government of Kenya [1986]).

Field interviews indicate that social pressure is a key mechanism for sustaining primary school contributions in rural western Kenya, and that ethnically diverse communities may be less able to effectively pressure parents to make school payments. The case of Matumbai Primary School in Teso district illustrates how low levels of inter-ethnic cooperation may lead to reduced educational investment. Matumbai is one of the most ethnically diverse schools in Teso district, with sizeable ethnic Luhya, Teso, and Kalenjin (a Nilo-Saharan group) communities. The headmaster of Matumbai primary school stated in June 2000 that ethnic "rivalry over ownership" of the school and over "who will take control of the school" was the central challenge facing Matumbai. Most parents have refused to fund *harambees* or participate in school meetings in Matumbai in recent years, due to a stated lack of trust across groups, and the absence of a feeling of "ownership" for the school. As a result, per pupil local school funding in 1995 was one-third of average local funding and no classrooms had been constructed at the school, so all classes took place under a tree – which meant that school was cancelled when it rained.[10]

Buduta Primary School in Busia district, a predominantly Luhya school with a substantial Luo minority, provides further evidence on how ethnic diversity may affect school fundraising. The headmaster of Buduta claims that he has had difficulties raising funds from the minority Luo community because they lack "a sense of ownership" for the school and "feel less committed to the school" than the majority Luhyas, although he stated that there are "no bad feelings [among Luhyas and Luos] in the school." The informal mechanisms that sustain Luhya school payments appear to be less effective among members of the minority Luo community, perhaps because they are outside the dense network of reciprocal social relations that sustain high contribution rates among the Luhyas (Woolcock [1998]).

In related work, Gugerty and Miguel [2000] examine the impact of ethnic diversity on Kenyan primary school committees, and find that higher diversity is associated with lower levels of community

[10] Due to recent financial assistance from the local NGO that provided the data for this study, two small temporary classrooms were finally constructed at Matumbai in 1999.

participation in school meetings, worse school committee attendance, and fewer school committee sanctions (recorded in meeting minutes) on parents who fail to make *harambee* contributions or pay fees.

### 2.3 Residential Mobility and School Transfers

Recent survey evidence indicates that land sales and residential mobility are rare in rural western Kenya, and this is true for a variety of reasons. First, land markets are thin in rural Kenya, as in much of Sub-Saharan Africa and many other poor countries (Collier and Gunning [1999]). Second, community approval is often required *de facto* to sell traditional family and kinship clan lands in rural Kenya, and this approval may be difficult to obtain from relatives (Platteau [1999]). Third, although relations between groups in Busia and Teso districts are peaceful, ethnic minorities are often treated with suspicion in this area as discussed above, discouraging residential mobility across traditional ethnic boundaries.

However, despite the low levels of residential mobility, survey evidence from 1996 indicates that school children in rural western Kenya are highly mobile by foot and frequently transfer among local primary schools: nearly half of the pupils in the sample have attended more than one primary school, 18 percent attend a school that is not the closest school to their home, and 14 percent reside and attend school in different geographic zones (local administrative units; there are 22 zones in Busia and Teso Districts).

## 3 A Theory of Ethnic Diversity and School Choice

The theoretical model explores how local school choice complicates the estimation of a causal relationship between ethnic diversity and school funding. The model examines school choice and funding decisions when: (i) mobility among schools is limited by land market imperfections; (ii) there is substantial exogenous variation in teacher quality; and, (iii) ethnic diversity directly affects the efficiency of educational provision. These three assumptions are meant to convey important aspects of school choice decisions in rural Africa that have not been examined in the existing literature, which is inspired by the United States case. The model presented in this section is related to Alesina, Baqir, and Easterly [1999], but incorporates sorting among schools as well as the above three features.

The principal theoretical result is that well-funded schools may attract more ethnically diverse pupil populations, if pupils from the surrounding area walk to good schools to take advantage of their educational quality. This suggests that the local correlation between funding and ethnic diversity among neighboring schools may often be positive, even if ethnic diversity has a negative impact on educational production. This sorting pattern also implies an upward bias in ordinary least squares estimates of the impact of ethnic diversity on school funding levels. These two implications of the theory are found to be consistent with the empirical results from western Kenya in section 4.

### 3.1 The Model Set-Up

The decision-making unit is a household. Each household belongs to one ethnic group, *A* or *B*. There is a continuum of households of unit measure evenly distributed along the unit interval ($x \in [0, 1]$), and one child from each household attends school. There are two schools, School 1 and School 2, located at $x = 0$ and $x = 1$, respectively. The number of schools is fixed.[11] The density of *A* households at location $x$ is $a(x) \in [0, 1]$, and the density of *B* households at location $x$ is $1 - a(x)$. Households choose the school that their child attends. Restricting the school choice decision to two neighboring schools reflects the underlying assumption of low household residential mobility.

Households receive exogenous income $y$. This is a one good economy, and households face a trade-off between private consumption $c$ and educational expenses $g$, such that $y = g + c$. Each household with a child in the school receives educational production $g^{\alpha}$ if all households pay $g$.[12] Households vote on the funding level in the school they have chosen for their child. Educational

[11] Founding a new primary school in Kenya is expensive, requiring the cost of classroom construction and the payment of teacher salaries for several years until the government recognizes it as an official school and begins to pay salaries. I assume for simplicity that this cost is large enough in the model that ethnic minority groups choose not to establish additional schools. Congestion effects from larger pupil populations are not considered in the model, although they could be introduced without changing the qualitative nature of the results.

[12] Primary education in Kenya has some characteristics of a private good since children whose parents do not pay fees or make contributions are at least temporarily suspended from school. However, it also has public good aspects since parents who do not participate in other important activities, including school meetings, are often not punished (Gugerty and Miguel [2000]).

production is concave in school funding ($\alpha \in [0,1]$) since basic investments like learning to read may have the highest payoffs.

Teacher quality, represented by $\varepsilon > 0$, also affects educational provision. $\varepsilon$ is given exogenously and reflects the motivation, competence, and honesty of the headmaster and teachers. The emphasis on exogenous elements of school quality distinguishes the model from existing school choice theories, and is appropriate in Kenya where all teachers and headmasters are centrally assigned to primary schools with minimal local input, as discussed above, and where the variation in headmaster quality across schools is thought to be large. The assumption that quality is at least partially observable to parents is realistic for Kenya, since average school scores on annual government examinations are publicly announced at local community meetings and are published in newspapers.

Ethnic diversity is allowed to directly affect the educational production technology. For example, more ethnic diversity may lead to lower levels of parent cooperation and participation in the school, more discord, and poor school committee performance. Alternatively, ethnic diversity may have a positive effect on educational outcomes by leading to the beneficial exchange of new ideas among people with different cultural backgrounds. Experimental social psychology evidence from the United States suggests that the ethnic diversity of work groups often affects group performance: ethnically diverse groups may be less productive than homogeneous groups at complex tasks requiring group consensus (Thomas [1999]), but they also produce more creative solutions to problems in some cases (McLeod, Lobel and Cox [1996]).[13] The direct effect of school ethnic diversity on the efficiency of educational production is represented in the model by the function $f_0(\gamma)$, where $\gamma$ is a measure of school ethnic diversity; $f_0(\gamma)$ may also be considered diversity's impact on the ability to collect school funds from parents.

[13] Thomas [1999] examines the performance of small groups (of three to four) university students in analyzing standard business school cases, while McLead, Lobel and Cox [1996] examine the performance of small groups of university students in proposing solutions to a business problem requiring creative brainstorming.

Effective educational production may be thought of as the discounted benefit of education in terms of future income. Taking into account teacher quality ($\varepsilon$) and school ethnic diversity ($\gamma$), effective educational production is:

*(1)* $$f_0(\gamma) \cdot \varepsilon \cdot g^{\alpha}$$

Equation 1 assumes that local school funding and teacher quality are complements in educational production: school funds are utilized more efficiently as teacher competence and honesty improve. This assumption seems appropriate in rural Kenya where schools with corrupt headmasters or high teacher absenteeism might squander other educational investments. Local funding and $f_0(\gamma)$ are also assumed to be complements, since parent may be unwilling to fund schools in which poor cooperation across ethnic groups paralyzes school committees.

Household utility is increasing in educational production and consumption, and is decreasing in travel costs to school. Walking to school imposes a cost proportional to the distance traveled to school, $|x - x_m|$, where $x_m$ is the location of School $m$ and $x$ is the location of the household on the unit interval. Utility for a pupil who lives at location $x \in [0,1]$ and attends School $m \in \{1,2\}$ is:

*(2)* $$U^x(m) = f_0(\gamma_m) \cdot \varepsilon_m \cdot g_m^{\alpha} + c - |x_m - x|$$

The timing is:

$t = 0$: Households observe school qualities ($\varepsilon_1$, $\varepsilon_2$).

$t = 1$: Households simultaneously make school choice decisions, and pay the sunk travel cost.[14]

$t = 2$: Majority voting in each school on the level of school fees ($g$) [15]

$t = 3$: Households receive income, pay fees, and receive education and private consumption.

[14] The assumption of a sunk travel cost is realistic in western Kenya, since neither the school fees nor the cost of a uniform particular to a school is refundable if the child transfers to another school during the school year.

[15] Attention is restricted to equilibria in which households vote sincerely, eliminating a class of equilibria generated by the fact that infinitesimal households' payoffs are invariant to their vote.

The model is solved using backwards induction, parameter values are common knowledge, and agents are forward-looking and have rational expectations. In $t = 2$, the majority group sets the level of school fees to maximize the utility of a representative majority group member taking into account school quality. The maximization problem faced by majority group households in $t = 2$ is:

$$g^* = \arg\max_g \left\{ f_0(\gamma) \cdot \varepsilon \cdot g^{\alpha} + (y - g) \right\} \tag{3}$$
$$= (\alpha \cdot f_0(\gamma) \cdot \varepsilon)^{\frac{1}{1-\alpha}}$$

Equation 3 implies is that parents are willing to spend more on school funding when educational investments are more productive, for example when the headmaster is more competent and honest and when there is more inter-ethnic cooperation.[16,17] Equation 3 can be re-written as 3′, where $f(\gamma) \equiv f_0(\gamma)^{\frac{1}{1-\alpha}}$, and $e \equiv (\alpha \cdot \varepsilon)^{\frac{1}{1-\alpha}}$, where $e$ can again be thought of as teacher quality.

$$g^* = f(\gamma) \cdot e \tag{3′}$$

Equation 3′ suggests the first-order approximation in Equation 4, which forms the basis for the empirical tests in section 4. $\tau$ denotes the marginal impact of a change in ethnic diversity on school funding. The estimated impact of ethnic diversity on school funding will be biased if schools' ethnic compositions are correlated with teacher quality, and if some aspect of teacher quality is unobserved to the researcher.[18]

$$g = a + \tau \cdot \gamma + \beta \cdot e \tag{4}$$

### 3.2 The Solution

[16] Although $g$ corresponds most closely to school fees in the context of Kenyan primary schools, Equation 3 may be understood as a reduced-form relationship between teacher quality and school funding appropriate in a variety of school finance settings. The theoretical results hold for community fundraisers (*harambees*) if contributions are also higher when educational investments are more productive.

[17] For simplicity, consider the case $y > g^*$: income levels are sufficiently high – alternatively, credit markets are sufficiently efficient – for households to afford the optimal primary school fee. However, the theoretical insights hold if this assumption is relaxed: at the corner solution $g_1 = g_2 = y$, effective educational production remains an increasing function of $\varepsilon$.

[18] If $e$ is unobserved in Equation 4, $\tau_{ols} = \tau + \beta \cdot [Cov(\gamma, e)/Var(\gamma)]$.

Subgame-perfect Nash equilibrium is the appropriate equilibrium concept for this dynamic game of complete information. *A* and *B* households residing at a given location *x* face the same maximization problem. Given common knowledge parameters $(\varepsilon_1, \varepsilon_2, \alpha)$ and the local density of Group *A* households ($a(x)$), a Nash equilibrium is characterized by a cut-off location $x^* \in [0, 1]$ such that all households with $x \leq x^*$ attend School 1 and all households with $x > x^*$ attend School 2. $\gamma_m$ is ethnic diversity in School $m \in \{1,2\}$ after school choice decisions, and is defined as the proportion of ethnic minorities in the school. When $\alpha = 0.5$ the following school choice indifference condition implies that the better funded school attracts more pupils, and draws pupils who live farther away on average:

*(5)* $$x^* = \frac{1}{2} + \frac{1}{2}\left(f(\gamma_1)\cdot e_1 - f(\gamma_2)\cdot e_2\right)$$

It is necessary to impose additional structure on the residential distribution of households in order to derive results for the relationship between school funding and school ethnic diversity, since there is no general relationship between equilibrium school ethnic diversity and school funding for all densities $a(x)$.[19] Attention is restricted to the linear function $a(x) = 1 - \phi \cdot x$, where $\phi \in [0, 1]$, such that the proportion of *A* households is weakly monotonically decreasing with distance from School 1. In ethnically diverse cases for $\phi$ near one, *A* households are numerically dominant in the area near School 1 and *B* households are dominant near School 2, while in ethnically homogeneous cases for $\phi$ near zero, *A* households are numerically dominant near both schools. Teacher quality shocks *e* are assumed to be distributed independently on the support $\{0, 1\}$ with $Pr(e = 1) = Pr(e = 0) = 0.5$ for simplicity, such that both schools are equally likely to have good teachers. Finally, the impact of ethnic diversity on educational production is represented by the function $f(\gamma) = \frac{1}{2} + d \cdot \gamma$, where $d \in [-1/2, 1/2]$. Diversity has a negative (positive) effect on educational production for $d < 0$ ($d > 0$).

[19] Result available from the author upon request.

**Proposition 1***:*

*(a) For high levels of ethnic diversity* $\bar{\phi} \leq \phi \leq 1$, *Corr*($g_s$ , $\gamma_s$) > 0 *among schools in the same geographic unit (s* $\in$ *{1, 2}).*

*(b) For low levels of ethnic diversity* $0 \leq \phi \leq \underline{\phi} < \bar{\phi}$ *Corr*($g_s$ , $\gamma_s$) *may be >0, <0, or =0 among schools in the same geographic unit (s* $\in$ *{1, 2}).*

*Proof: Refer to the Theory Appendix.*

Proposition 1 applies to neighboring schools within a single geographic "unit". Proposition 1(a) implies that school funding ($g$) and ethnic diversity ($\gamma$) are positively correlated across neighboring schools in the ethnically diverse case (for high $\phi$). The better-funded school attracts an ethnically diverse pupil population due to its superior educational quality, while the other school is poorly funded, of low educational quality, is relatively ethnically homogeneous, and has a smaller pupil population. This relationship holds even when greater ethnic diversity is associated with lower educational production ($f' < 0$): the better quality school becomes relatively less attractive as it becomes more ethnically diverse, dampening the magnitude of pupil flows but leaving the pattern of ethnic diversity and funding across neighboring schools intact. It is straightforward to prove that Proposition 1(a) also holds in the case of complete ethnic segregation and equal-sized groups ($a(x) = 1$ for $0 < x < 1/2$, $a(x) = 0$ for $1/2 < x < 1$).

Proposition 1(b) indicates that there is no general relationship between equilibrium ethnic diversity and school funding levels in the ethnically homogeneous case (for low $\phi$). In the case of $\phi = 0$, it is trivial to see that funding and diversity are uncorrelated across neighboring schools since ethnic diversity is always zero in both schools. When there are a minority of *B* households residing near School 2 ($\phi$ greater than zero, but small) a positive school quality shock in School 2 leads a larger proportion of *A* households (who live farther away) to attend the school, making School 2 less ethnically diverse; however, a positive quality shock in School 1 leads a larger proportion of *B* households to attend the school, making School 1 more diverse. Proposition 1(b) indicates that the correlation between funding and diversity among neighboring schools may in general be positive, negative or zero in this case because of these two opposing effects.

### 3.3 Multiple Geographic Units

The model can be extended to include multiple geographic units. The assumption is maintained throughout that there is no pupil mobility between geographic units, only within units, reflecting the low levels of residential mobility in rural western Kenya. The relationship between funding and ethnic diversity across all schools when there are many geographic units reflects both the local relationship between diversity and funding due to pupil sorting *within* units (Proposition 1), and the impact of ethnic diversity on funding *across* units, where the latter effect depends on the sign of $f'$. The cross-sectional relationship between local school funding and school ethnic diversity across schools that are not within walking distance of each other may be either positive or negative due to these two potentially offsetting effects. For concreteness, there are measure one geographic units indexed $u \in [0, 1]$ where a fraction $D$ of the units are high ethnic diversity units ($\phi = 1$) and a fraction $1 - D$ are low ethnic diversity units ($\phi = 0$). A higher proportion of diverse units (higher $D$) implies greater overall ethnic diversity in the area.

Proposition 2 presents the remaining theoretical results. $g_{su}$ represents school funding for school $s \in \{1, 2\}$ in area $u$, $\gamma_{su}$ is ethnic diversity for school $s$ in area $u$, $\overline{g_u}$ is average school funding for the two schools in area $u$, and $\overline{\gamma_u}$ is average ethnic diversity in area $u$. Proposition 2(a) indicates that the correlation between ethnic diversity and school funding across schools is greater than the correlation across area averages, due to the positive local sorting effect in high diversity areas described in Proposition 1(a). Proposition 2(b) implies that the cross-sectional correlation between school funding and school ethnic diversity – across all schools in the multiple geographic units – may be either positive or negative: when the proportion of high ethnic diversity units is sufficiently large, this correlation is positive, reflecting the positive local sorting effect identified in Proposition 1(a). Proposition 2(c) indicates that the correlation between average school funding and average ethnic diversity across geographic units is negative (positive) when ethnic diversity has a negative (positive) effect on educational production.

**Proposition 2**:

*(a)* $Corr(g_{su}, \gamma_{su}) > Corr(\overline{g_u}, \overline{\gamma_u})$ *, where* $s \in \{1, 2\}$ *and* $u \in [0, 1]$.

*(b)* $Corr(g_{su}, \gamma_{su}) > 0 \Leftrightarrow D > 1 - \frac{4(d+2)^2}{d(d-4)(10-d)}$ *, where* $s \in \{1, 2\}$ *and* $u \in [0, 1]$.

*(c)* $Corr(\overline{g_u}, \overline{\gamma_u}) < 0 \Leftrightarrow d < 0$ *, where* $u \in [0, 1]$.

*Proofs: Refer to the Theory Appendix.*

### 3.4 Empirical Implications of the Theory

The following empirical tests examine whether pupil sorting among primary schools in western Kenya is consistent with the theory. $\tau$ denotes the true marginal effect of ethnic diversity on school funding and $\tau_{ols}$ denotes the ordinary least squares regression estimate of this effect (regression 4).

(i) *Test of Proposition 1(a): Corr(school ethnic diversity, local school funding) > 0 among neighboring schools in ethnically diverse areas*

Proposition 1(a) implies that school ethnic diversity is *positively* correlated with local school funding among neighboring schools in ethnically diverse areas, as well-funded and high quality schools attract diverse pupil populations. The local relationships between school funding and school ethnic diversity; average school exam performance and school ethnic diversity; and total pupil enrollment and school ethnic diversity are estimated in section 4 using an OLS framework with geographic zone fixed effects.

(ii) *Test of Proposition 2(a):* $\tau_{ols} > \tau$

Proposition 2(a) and Equation 4 suggest that the estimated OLS coefficient on ethnic diversity suffers from upward omitted variable bias when school ethnic diversity and school funding are both positively related to unobserved teacher quality. This implication is tested in western Kenya by comparing ordinary least squares and instrumental variable estimates of the impact of ethnic diversity on school funding. The

instrumental variable for endogenous school ethnic diversity is local residential ethnic diversity, avoiding the sorting effects that may bias OLS estimates.

## 4 Empirical Framework and Results

### 4.1 Data

Detailed data for one hundred of the three hundred thirty-seven rural primary schools in Busia and Teso districts were collected from pupil, school, and teacher questionnaires in early 1996 as baseline information for a non-governmental organization (ICS Africa) School Assistance Project (SAP). The schools were selected by the Ministry of Education district education office in 1995 using the following criterion (Glewwe, Kremer, and Moulin [1998]): these were needy rural primary schools that had not been assisted by an earlier World Bank assistance program, which had distributed free textbooks to one hundred different schools in this area in 1994. The SAP schools were first stratified by geographic division and then randomly assigned into four groups that were to receive financial assistance in sequence over the following four years. Table 1 indicates that the School Assistance Project (SAP) schools had smaller enrollments and performed worse on government exams than other rural schools in the area, due to the selection criteria. However, local ethnic diversity at the level of the geographic zone – a local administrative unit – is not significantly associated with additional differences between the SAP and non-SAP schools in terms of enrollment and exams, suggesting that sample selection is unlikely to be producing a spurious correlation between local ethnic diversity and school outcomes in this sample.

The non-governmental organization collected a variety of financial and demographic data for these schools in 1996. The pupil questionnaire focused on pupils' schooling background, family educational characteristics and asset ownership, and self-described ethnic affiliation. Over six thousand pupil questionnaires were administered by trained survey enumerators to all grade six through eight pupils (ages twelve to eighteen years) present on the day of questionnaire administration in 1996. Younger pupils were not administered questionnaires because of their limited reading and writing skills. In total, 100 school questionnaires and 861 teacher questionnaires were also administered in early 1996. School

questionnaires – filled by schoolmasters with the assistance of a trained enumerator – contain detailed information on school finances, infrastructure, inputs, and pupil enrollment. Teacher questionnaires focus on teacher qualifications, and were completed by the teachers themselves.[20] The Busia District Education Office provided school examination results and exam name lists. NGO enumerators using portable Global Positioning System (GPS) machines collected school latitude and longitude. The analysis below includes the ninety primary schools with complete pupil, school, teacher, and GPS data.[21]

For comparability with the existing literature (Mauro [1995], Easterly and Levine [1997] and Alesina et al [1999]), ethno-linguistic fractionalization (ELF) is used as the principal empirical measure of ethnic diversity. Ethno-linguistic fractionalization is the probability that two people randomly drawn from the population are from distinct groups. Formally,

*(6)* $$ELF \equiv 1 - \sum_i (\text{Proportion of Ethno-linguistic group } i \text{ in the population})^2$$

Using the 1996 Pupil Questionnaire data, ethno-linguistic fractionalization among all sample pupils is 0.48 when Luhya subtribes are considered a single ethnic group. By way of contrast, median ELF among U.S. counties in 1994 was 0.14 (Alesina et al [1999]), considerably less diverse than Busia and Teso districts. Additional measures of ethnic diversity employed in the analysis include ethno-linguistic fractionalization among the Luhya subtribes, and one minus the proportion of the largest ethnic group in a school (aggregating Luhya subtribes).

The main outcome is total local school funding collected per pupil in 1995. This does not include funds raised from non-governmental organizations; however, since schools that were receiving considerable outside donor assistance were largely excluded from the program, only six sample schools had received over $100 in outside funding in 1995 and local fundraising does not appear to be

[20] The survey enumerators – Charles Asoka, Robert Namunyu, Polycarp Waswa, and Maureen Wechuli – believe that responses from the school and teacher questionnaires are more reliable than pupil responses. Fortunately, the question on pupil ethnic affiliation is likely to suffer from less response error than other questions.

[21] These data are available from the author upon request.

significantly crowded out in these schools (regressions not shown). School fees collected per pupil and total local *harambee* donations per pupil are additional school finance measures. School facilities - the number of desks per pupil, latrines per pupil, and classrooms per pupil in 1996 - are also considered as outcome measures since chronically under-funded schools are likely to have fewer educational resources than other schools. The number of school-owned textbooks per pupil is another measure of past local educational investment. Average school performance on NGO academic examinations (which were based on the format of government exams) for grades three to eight in 1996 captures aspects of educational quality.

Tables 2 and 3 present descriptive statistics and suggest that there is considerable variation across schools in ethnic diversity, and in local funding levels and other outcome measures. Table 4 indicates that observed socioeconomic differences across ethnic groups are small and not systematic, suggesting that ethnic diversity is unlikely to be proxying for income inequality or average socioeconomic status. [22]

**4.2 Identification Strategy**

The exogeneity of ethnic land settlement patterns in Busia and Teso districts forms the basis for the empirical identification strategy employed in this paper. A variety of evidence suggests that current levels of local ethnic diversity in Busia and Teso districts are largely the product of historical accident rather than recent migration.[23] "The nineteenth century was a time of considerable unrest throughout the District, with conflict between the Luhya groups, Luo, Teso and Kalenjin" (Government of Kenya [1986]). Were [1967] writes that "various factors - famine, epidemics, domestic disputes, the spirit of adventure and warfare - made the inhabitants of the region extremely mobile" (p. 41) from the 17th century through the 19th century, when various Nilo-Saharan ethnic groups migrated to western Kenya

[22] Unfortunately, there is insufficient information on household incomes, consumption, and land ownership in the dataset to directly examine the relationship between funding outcomes and income inequality.

[23] Religious diversity is not included as an explanatory variable in the empirical analysis, since local religious affiliation is not plausibly exogenous due to the extensive missionary activity in this area during the past century. A negative correlation between religious fragmentation and school funding cannot be interpreted as a causal relationship if evangelical activity is targeted to and is most successful in poor areas, for example.

from present-day Uganda. Successive waves of Teso and Luo migration, and the resulting wars with established Luhya communities, largely determined ethnic residential patterns in the area.

The emergence of British colonial authority in western Kenya in 1894 ended wars and cattle raiding, and the population movements that accompanied them. Morgan et al. [1966] write that ethnic land claims were "frozen by the Colonial Government by the demarcation of 'African Land Units'. This prevented the expansion of tribes into another's territory and thus eliminated the principal source of major inter-tribal wars. ... Within the African areas the indigenous pattern of 'water-tight' units was maintained, but accentuated by the increasing population." Land demarcation and individual land registration during the post-colonial period "has frozen the previously fluid situation and virtually halted the traditional mobility" (Government of Kenya [1986]). Unlike the central highlands of Kenya, Busia and Teso were largely free from European settlement - and resulting disruptions of ethnic land claims - during the colonial period.

Figure 2 presents a map of Busia and Teso districts, including geographic division boundaries and the location of all sample (SAP) and non-sample schools. Kenya is divided into 50 districts, each of which is split into divisions, which are in turn composed of geographic zones. There are two principal measures of local ethnic diversity employed in this study. The ethnic diversity of pupils residing in a school's geographic zone – a measure of diversity independent of local pupil sorting among schools – is used as an instrumental variable for school-level diversity to address the sorting bias characterized in the theory. Zonal ethnic diversity is computed among all surveyed pupils from sample schools residing in the corresponding zone using the 1996 Pupil Questionnaire data.[24] When indicator variables for geographic divisions are included to control for regional differences in income and tastes for educational spending, the relationship between diversity and school outcomes is identified across zones within each division.

[24] Although the pupil questionnaire data do not contain ethnic affiliation information for the youngest pupils (below grade 6), they do indicate that drop-out rates are similar across ethnic groups in grades 6 to 8, suggesting that differential school participation across groups is unlikely to significantly alter measured ethnic diversity.

A second measure of local ethnic diversity is computed among all pupils in primary schools located within five kilometers of each school. These data were created from official 1996 government examination name lists provided by the Busia and Teso District Education Office. Portable global positioning system (GPS) technology was employed to determine the latitude and longitude of primary schools in Busia and Teso Districts. The five kilometer radius around each school appears to be a plausible upper limit on the distance that children may walk to school on a daily basis. The results in Tables 8 and 9 are robust to radius distances of between three to six kilometers (results not shown).

The principal advantage of the data constructed from government examination name lists is that it includes information for nearly all primary schools in Busia and Teso districts (326 of 337 schools), limiting possible bias from the selection of schools into the NGO assistance program in 1996 – although Table 1 suggests that this bias is unlikely to be large. However, a drawback of these data is that pupil ethnic affiliation was assigned by NGO staff based upon children's names rather than being determined by pupils themselves.[25] The assignment of ethnicity by NGO staff is likely to introduce considerable error into measured school ethnic diversity, since many surnames are common across ethnic groups in this area, and names and ethnic affiliation often do not match up. It is particularly difficult to distinguish between Luhya and Luo children since many Luhyas possess Luo surnames: approximately nineteen percent of all pupils in the exam name list sample have ambiguous Luo surnames. Pupils with ambiguous names are assigned Luhya and Luo ethnicity in proportion to their group's representation within the geographic zone in the 1996 Pupil Questionnaire sample; this means that pupils with ambiguous names are more likely to be assigned Luo ethnicity in areas in which the Pupil Questionnaire data indicate that there are more Luos. Despite the possible error inherent in assigning pupil ethnicity based on their name, the two measures of local ethnic diversity are quite highly correlated (correlation coefficient 0.7).

Table 5 examines the relative stability of residential ethnic composition in Busia and Teso districts during the post-colonial period as a test of the validity of the identification strategy. Comparing

[25] Charles Asoka and Maureen Wechuli assigned pupil ethnicity to the name list data.

residential ethnic composition at the geographic division level in 1996 using ICS Pupil Questionnaire data to residential composition in 1962 using Kenyan Census data suggests that ethnic residence patterns have been largely stable: the ordering of residential ethnic diversity across geographic divisions, measured either by the size of the largest ethnic group, is identical in 1962 and 1996. Unfortunately, census data on ethnic composition at the geographic zone level is unavailable. Recent survey evidence also suggests that land sales and residential mobility are extremely rare in Busia and Teso districts: among 507 local households interviewed for a different study in this area, only three respondents claimed to have bought or sold any land during 1997 or 1998, and these individuals may have purchased land in the vicinity rather than moving away (Gugerty and Miguel [2000]).

Children may move in with relatives to attend a primary school that is not within walking distance of their home, altering effective local ethnic composition even if adult residential patterns are fixed. Table 4 presents evidence that fewer than 15 percent of pupils are not living with a parent, among pupils with at least one surviving parent. Since some pupils move in with relatives residing in the same geographic zone - relatives often live near each other - the proportion of children who move in with relatives in a different geographic zone is plausibly considerably less than 15 percent, though data limitations make it impossible to determine the exact proportion. Such rates are unlikely to significantly alter local residential ethnic diversity.

Table 6 presents the first stage regressions of local ethnic diversity (explanatory variable) on school ethnic diversity (dependent variable), and indicates that both zonal ethnic composition and ethnic composition within a five kilometer radius are strong predictors of school ethnic composition: local ethnic diversity alone captures over 40 percent of the variation in school-level ethnic diversity.

Table 7 presents the reduced-form regressions of local ethnic diversity on local school funding (dependent variable), and indicates that higher local residential ethnic diversity is associated with significantly lower local school funding. However, in addition to any direct impact it may have on local collective action, ethnic diversity may be associated with local school funding through its relationship with other local characteristics; for example, ethnically diverse regions may be poorer than other areas

because it is difficult to enforce contracts within heterogeneous communities, leading credit, land and labor markets to function less efficiently there (Besley and Coate [1995]; Greif [1993]). Ethnically diverse areas may also have worse schools if they are assigned lower quality teachers by the national Ministry of Education.

Table 7 presents the reduced-form relationships between the instrumental variables and a variety of observed socioeconomic and school quality measures, including average father's education and formal sector employment, latrine ownership at home, fertility, years since school founding, total pupil enrollment, pupil-teacher ratio, as well as teacher and headmaster qualifications, gender, and experience. With the exception of latrine ownership (which is considered an indication of higher socioeconomic status and is positively and significantly associated with local ethnic diversity in both specifications) and perhaps father's formal sector employment (which is negatively associated with local diversity in one specification) none of these measures is significantly associated with local ethnic diversity and there is no systematic pattern of coefficient estimates, suggesting that differences in socioeconomic and teacher characteristics across areas are unlikely to be driving the pattern between diversity and school funding.

The second stage regression is presented in Equation 7. To control for socioeconomic and cultural variation across geographic areas, average fathers' education, average latrine ownership, the average number of siblings among pupils residing in the geographic zone, indicator variables for geographic divisions, teacher characteristics, as well as the proportions of various ethnic groups in the geographic zone – which may capture average differences in the taste for education across different ethnic groups – are included as explanatory variables in some specifications. $Y$ is the outcome measure - such as school funding - $E$ is school-level pupil ethnic diversity (instrumented with local ethnic diversity in the first stage), $G^j$ are $j$ geographic division indicator variables, $Z^k$ are $k$ zonal socioeconomic, demographic, and teacher controls, and $i$ denotes a school. In the specifications using zonal residential ethnic diversity as an instrumental variable for school ethnic diversity, school regression disturbance terms are assumed to be independent across geographic zones but are allowed to be clustered within geographic zones. When local ethnic diversity among schools within five kilometers of the school is used as the instrumental

variable for school ethnic diversity, regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance, using the spatial generalized method of moments (GMM) estimation method in Conley [1999].[26]

*(7)* $$Y_i = a + \pi E_i + \sum_j \beta^j G_i^j + \sum_k \delta^k Z_i^k + \mu_i$$

**4.3 Empirical Results**

Table 8 presents the main empirical results. There is an insignificant negative relationship between ethno-linguistic fractionalization across tribes and school funding in the ordinary least squares specification in Regression 1. Figure 3 presents the cross-sectional relationship between school funding and school-level diversity, illustrating the weak relationship. However, the instrumental variable coefficient point estimates on ELF are negative and significantly different than zero at 95 percent confidence (Regression 2). This relationship is robust to the inclusion of geographic division indicator variables (Regression 3), zonal socioeconomic controls (Regression 4), and the proportion of each ethnic group in the zone as well as teacher characteristics (Regression 5), suggesting that measured ethnic diversity is not proxying for average socioeconomic status, cultural differences, or teacher quality across areas. Figure 4 graphically presents the negative relationship between average school funding and residential ELF across geographic zones. There is a sharp reduction in the absolute magnitude of the coefficient estimate on ELF when geographic division indicators are included in Regression 3, indicating that the relationship between ethnic diversity and school funding across geographic divisions accounts for much of the overall negative relationship.

An interpretation of the instrumental variable coefficient estimate in Regression 4 is that the drop in local school funding associated with a change from complete ethnic homogeneity to median school-level ethnic diversity is 31 shillings, or nearly 25 percent of average local funding. Since an average

[26] Following Conley [1999], spatial standard errors are calculated with a weighting function that is the product of a kernel in each direction (North to South, East to West); the kernels start at one and decrease linearly until they are zero at 8 km from the school; results are robust to varying this cut-off between 5 to 8 km (results not shown).

primary school textbook costs approximately 150 Shillings, and the ratio of textbooks to pupils in these schools is one to three (Table 2), the result suggests that eliminating the "costs" associated with higher ethnic diversity would allow diverse schools to more than double their textbook stocks over the course of two years.

In Regression 6, the coefficient estimate on ethnic diversity across tribes is negative and significantly different than zero at 99 percent confidence, while the coefficient estimate on ethnic diversity among Luhya subtribes is insignificantly different than zero. This suggests that ethnic diversity across tribes – rather than across Luhya subtribes – accounts for the observed negative relationship between ethnic diversity and school funding.[27]

ELF could be capturing a non-linearity in the relationship between funding and the size of a particular group rather than the impact of ethnic diversity *per se*. However, Regression 7 provides evidence that the functional form of the ethno-linguistic fractionalization index is not driving the results: a linear measure of ethnic diversity - the proportion of the largest ethnic group in the school - is negatively and significantly related to the level of local school funding per pupil. As in Regression 5, the relationship between the proportion of the largest ethnic group in the school and local funding is robust to the inclusion of the proportions of various ethnic groups as explanatory variables (results not shown). Figure 5 indicates that ethnic diversity is not proxying for the proportion of Tesos in the area: among the predominantly ethnically Teso geographic zones, more ethnically diverse zones have lower average funding than homogeneous zones.

Regression 8 presents the relationship between school ethnic diversity and local funding, using ethnic diversity among all schools located within five kilometers of the school as the measure of local ethnic diversity. Standard errors are corrected to allow regression disturbance terms to be correlated across schools as a function of their physical distance (Conley [1999]). The results suggest that school

[27] Field interviews conducted by the author suggest that disputes between kinship clan lines (from the same ethnic group) sometimes occur in these schools. Unfortunately, there is insufficient information on clan affiliation to explore the impact on school funding.

ethnic diversity is strongly associated with lower school funding, and the coefficient estimate on ethnic diversity is significantly different than zero at 95 percent confidence. The point estimate is similar in magnitude to the analogous coefficient in Regression 2, suggesting that the results are robust to this alternative source of ethnicity data. Regression 9 presents the results including geographic zone socioeconomic controls, and also yields a large and statistically significant (at 90 percent confidence) negative estimate of the relationship between diversity and school funding. However, when geographic division indicators are included in Regression 10, the point estimate on school ELF remains negative but is no longer statistically significantly different than zero. This result again indicates that the relationship across geographic divisions accounts for much of the overall negative relationship between ethnic diversity and local school funding.

The regressions in Table 9, which include geographic zone socioeconomic controls as explanatory variables, indicate that ethnic diversity is significantly negatively associated with donations from community fundraisers (*harambees*), but not significantly associated with school fees collected per pupil. This implies that lower levels of voluntary *harambee* contributions in diverse areas accounts for most of the decline in school funding. *Harambees* are occasions when community members without children in the school as well as local politicians – in addition to parents of school children – have the opportunity to publicly contribute to a school project. This result suggests that the school committee may be able to more effectively encourage parents to contribute in ethnically homogeneous schools; that residents of ethnically homogeneous areas may feel more "ownership" for their local schools than residents of diverse areas; or that politicians tend to assist their ethnically homogeneous political strongholds. Unfortunately, it is impossible to distinguish between these hypotheses since the dataset does not contain the proportions of *harambee* contributions that come from parents, other community residents, and local politicians.

Table 9 also presents the relationship between ethnic diversity and primary school facilities, which reflect the cumulative impact of past educational investments as well as past outside assistance. The coefficient estimates on ethnic diversity are large, negative and significantly different than zero for

desks per pupil: the drop in desks per pupil associated with a change from complete ethnic homogeneity to median school-level ethnic diversity is over 25 percent of average desks per pupil. In the regressions with classrooms per pupil and pupil latrines per pupil regressions as dependent variables, the coefficient estimates on ethnic diversity are large and negative in all specifications, although insignificantly different than zero at traditional confidence levels. Taken together, these results suggest that primary schools in ethnically diverse areas may have worse facilities than schools in more homogeneous areas. In addition to their impact on learning, infrastructure investments directly enhance pupil utility; classrooms with a sturdy roof shield children from rain, and latrine construction is an important public health intervention.

Table 9 reports the relationship between local ethnic diversity and the density of primary schools, and finds that there are significantly fewer primary schools in diverse areas, perhaps due to collective action problems with regard to setting up schools in these areas: zones with average levels of local ethnic diversity contain approximately twenty percent fewer schools than ethnically homogeneous areas. However, Table 7 indicates that local ethnic diversity is not significantly associated with total pupil enrollment, implying that the lower density of schools has not increased school crowding in diverse areas. Taken together, these results suggest that ethnically diverse areas are either less densely populated on average, or that school enrollment rates are lower in ethnically diverse areas, perhaps as a result of the lower density of schools; the data do not permit ruling out either of these two possibilities.

Table 9 also examines the stock of school textbooks per pupil as the dependent variable and indicates that the coefficient estimates on ethnic diversity are negative but not significantly different than zero. The relationship between ethnic diversity and the number of privately owned textbooks per pupil is also reported in Table 9 to explore the possibility of substitution from publicly provided to privately-owned textbooks in ethnically diverse areas. The coefficient point estimates on ethnic diversity are positive in this case but not statistically significantly different than zero. The result serves as a specification check, suggesting that unobserved differences in the taste for education or income across areas – which would affect private textbook ownership as well as public funding outcomes – are unlikely to be driving the relationship between ethnic diversity and local school funding.

Finally, Table 9 reports the relationship between ethnic diversity and average school scores on NGO examinations for pupils for grades three to eight in 1996. The coefficient estimates are nearly zero in both specifications, which may be surprising in light of the negative relationship between ethnic diversity and local funding. However, other recent studies from rural western Kenya have found that average school exam scores respond little to increases in educational inputs, including textbooks, classroom construction, and school health programs (Glewwe, Kremer, and Moulin [1998]; Miguel and Kremer [2000]). This result is not necessarily inconsistent with the theoretical model in Section 3, in which ethnic diversity affects educational production: ethnic diversity may affect aspects of educational production not captured on government exams, and ethnic diversity is associated with worse school facilities (Table 9), which may directly affect the quality of pupils' educational experience.

**4.4 Testing the Theory**

*Test of Proposition 1(a)*

Table 10 estimates an OLS regression of school funding (dependent variable) on school ethnic diversity (explanatory variable) including geographic zone fixed effects. The inclusion of zonal fixed effects isolates the local relationship between ethnic diversity and school funding among schools within the same geographic zone. The average geographic zone in Busia and Teso districts is eighty $km^2$ in area which corresponds nearly exactly to the area of the five kilometer radius neighborhoods used in the spatial GMM specifications and seems a reasonable bound on pupils' walking radius.

The theory predicts that the coefficient on ethnic diversity will be positive within ethnically diverse areas – which includes much of Busia and Teso districts (Figure 4) – as good quality schools become both more ethnically diverse and better funded than neighboring schools (Proposition 1(a)), while the relationship in less diverse areas is theoretically ambiguous (Proposition 1(b)). The coefficient estimates on school ELF are indeed positive in Table 10 for both total local school funds and total donations per pupil, and the coefficient estimate is positive and significantly different than zero at 90 percent confidence for total school fees collected per pupil. Table 10 also examines the local relationship

between funding and ethnic diversity for geographic zones with above average ethnic diversity (zonal ELF greater than 0.218) and for zones with below average ethnic diversity, and finds that the coefficient point estimate on ethnic diversity among the more diverse zones is substantially larger than the estimate for the less diverse zones (101.8 to 63.6). This pattern of coefficient estimates is consistent with Proposition 1.

Proposition 1(a) also implies that better quality schools become more ethnically diverse in high ethnic diversity areas, and this is confirmed in the data: the point estimate on school ELF is positive and significantly different than zero at 95 percent confidence when the average school score on 1996 NGO exams is the dependent variable. The coefficient estimate on school ELF among geographic zones with above average ethnic diversity is also positive and significantly different than zero at 95 percent confidence (coefficient estimate 1.01), while the coefficient estimate on school ELF among geographic zones with below average ethnic diversity is near zero (0.09).

Table 10 also presents the OLS regression with zonal fixed effects for total pupil enrollment in the school in 1996 as the dependent variable, and indicates that the point estimate on school ethnic diversity is positive in this case. Although the coefficient estimate is not significantly different than zero, this is further suggestive evidence in favor of the theoretical sorting model.

*Test of Proposition 2(a)*

The estimated ordinary least squares coefficient on ethnic diversity (-22.7) is greater than the instrumental variable estimate (-191.5) in Table 8, which is consistent with the theoretical model. However, the pattern of coefficient estimates may also be caused by attenuation bias due to measurement error in school-level ethnic diversity. Measurement error is likely since ethnic affiliation information from the pupil questionnaire is available for a subsample of each school: on average, pupil questionnaires are available for only 68 pupils per school (total average enrollment is 296 pupils). For attenuation bias to

fully account for the difference between the IV and OLS coefficient estimates, however, measurement error would have to constitute an implausibly high 88 percent of the variation in school-level ELF.[28]

## 5 Conclusion

To summarize the empirical findings, ethnic diversity is associated with sharply lower local school funding and worse school facilities in rural western Kenyan primary schools. Ethnic diversity across tribes rather than across subtribes appears to be driving this negative relationship, suggesting that collective action problems may be more severe in the presence of greater cultural and linguistic differences. Donations from local fundraisers (*harambees*) – events that require considerable community coordination and participation to be successful – are sharply lower in ethnically diverse areas. Given the central role that *harambees* play in Kenyan local public finance, the results of this paper suggest that local ethnic diversity may be negatively associated with the provision of other local public goods in Kenya – a hypothesis that the author will test in future research. Finally, local pupil sorting among schools in western Kenya is consistent with the theoretical prediction that good quality schools tend to attract an ethnically diverse pupil population from the surrounding area.

There is no obvious policy prescription for the negative ethnic diversity effects described in this paper. Governments could subsidize the creation of additional primary schools in diverse areas to facilitate sorting into ethnically homogeneous schools and thereby avoid the efficiency cost of diversity. However, promoting ethnic separatism may have deleterious long-term political implications in sub-Saharan Africa, where ethnic divisions have often been associated with violent conflict. On the other

[28] The OLS coefficient in the presence of attenuation bias is $\beta_{OLS} = \beta \cdot (\sigma^2_{ELF} / [\sigma^2_{ELF} + \sigma^2_u])$, where $\beta$ is the true coefficient on ELF, $\sigma^2_{ELF}$ is the variance of school ELF, and $\sigma^2_u$ is the variance of measurement error. Simulations suggest that attenuation bias due to sampling variation in the pupil questionnaire sample should account for less than 15 percent of the difference between the IV and OLS estimates (results not shown).

hand, further centralization of school funding could increase regional and ethnic favoritism in the allocation of national government funds, which is often extreme in Kenya and other African countries.[29]

A more attractive approach for addressing the efficiency costs of ethnic diversity lies in identifying the mechanisms through which diversity affects organizational performance, in order to design policies and institutions that promote successful collective action. The author's field work for this project, as well as related work in Gugerty and Miguel [2000], points to the important role that community social sanctions may play in sustaining local public goods provision in less developed countries, mechanisms that are most effectively applied *within* social groups. A long-standing theme among observers of economic development is that the formation of meaningful economic linkages extending beyond the immediate community is a necessary pre-condition for modern economic growth (Simmel [1971 (1908)]; Greif [1993]; Woolcock [1998]). The design of policies that build cooperation, or "social capital" (Putnam [1993]), *across* ethnic groups – perhaps including central government nation-building efforts, or power sharing arrangements within organizations – remains a poorly understood yet promising research agenda with critically important implications for economic development in sub-Saharan Africa and elsewhere. [30]

[29] Barkan and Chege [1989] study the allocation of national road construction funds in Kenya during the 1970s and 1980s, and find that the proportion of road funds allocated to the ethnic homeland of former Kenyan President Jomo Kenyatta fell from 44 percent in 1979-1980 to 16 percent in 1987-1988 after Kenyatta's Kikuyu ethnic group lost its dominant position in the central government, while the ethnic homeland of Kenyan President Daniel Arap Moi – who replaced Kenyatta – saw its share of road funds rise from 32 to 57 percent during the same period.

[30] Horowitz [1985] is the seminal discussion of ethnic conflict, and Carroll and Carroll [2000] review the current state of this literature. Hawley [1981] concludes that administrators in successfully desegregated schools in the U.S. "ensure that persons of different races share positions of status and power within the school". Fearon and Laitin [1996] model ethnic cooperation as a matching game in which each ethnic group disciplines its own members if they transgress a reciprocity norm. Barkan [1994] discusses the serious nation-building efforts in Tanzania since independence.

**6 References**

Alesina, Alberto, Reza Baqir, and William Easterly. (1999). "Public goods and ethnic divisions", *Quarterly Journal of Economics*, 114(4), 1243-1284.

Alesina, Alberto, and Eliana Laferrara. (2000) "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, 115(3), 847-904.

Barkan, Joel D. (1994). "Divergence and Convergence in Kenya and Tanzania: Pressures for Reform," in *Beyond Capitalism versus Socialism in Kenya and Tanzania*, ed. Joel D. Barkan. Lynne Rienner Publishers: Boulder.

Barkan, Joel D., and Michael Chege. (1989). "Decentralising the state: district focus and the politics of reallocation in Kenya", *Journal of Modern African Studies*, 27(3), 431-453.

Benabou, Roland. (1993). "Workings of a city: location, education, and production", *Quarterly Journal of Economics*, 108 (3), 619-652.

Besley, Timothy, and Stephen Coate. (1995). "Group lending, repayment incentives and social collateral", *Journal of Development Economics*, 46(1), 1-18.

Carroll, Barbara Wake, and Terrance Carroll. (2000). "Accommodating ethnic diversity in a modernizing democratic state: theory and practice in the case of Mauritius," *Ethnic and Racial Studies* 23(1), 120-142.

Collier, Paul, and Jan Willem Gunning. (1999). "Explaining African economic performance," *Journal of Economic Literature*, 37 (1), 64-111.

Conley, Timothy. (1999). "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 92(1), 1-45.

Easterly, William, and Ross Levine. (1997). "Africa's growth tragedy: policies and ethnic divisions", *Quarterly Journal of Economics*, 112 (4), 1203-1250.

Fearon, James, and David Laitin. (1996). "Explaining Interethnic Cooperation", *American Political Science Review*, 90(4), 715-735.

Fernandez, Raquel and Richard Rogerson. (1996). "Income distribution, communities, and the quality of public education", *Quarterly Journal of Economics*, 111(1), 135-164.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. (1998). "Textbooks and test scores: Evidence from a prospective evaluation in Kenya", mimeo, MIT.

Goldin, Claudia and Lawrence F. Katz. (1997). "Why the United States led in education: Lessons from secondary school expansion, 1910 to 1940", NBER Working Paper #6144.

Government of Kenya. (1965). *Kenya Population Census, 1962*. Central Bureau of Statistics: Nairobi.

Government of Kenya. (1986). *Kenya Socio-Cultural Profiles: Busia District*, ed. Gideon Were. Ministry of Planning and National Development, Nairobi.

Government of Kenya. (1994). *Kenya Population Census, 1989*. Central Bureau of Statistics: Nairobi.

Greif, Avner. (1993). "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition", *American Economic Review*, 83(3), 525-548.

Gugerty, Mary Kay, and Edward Miguel. (2000). "Community Participation and Social Sanctions in Kenyan Schools" mimeo., Harvard University.

Hawley, Willis. (1981). "Equity and Quality in Education: Characteristics of Effective Desegregated Schools", in *Effective School Desegregation: Equity, Quality and Feasibility*, ed. Willis Hawley. Sage Publications: London.

Horowitz, Donald L. (1985). *Ethnic Groups in Conflict*. University of California Press: Berkeley.

Krueger, Alan B., and Mikael Lindahl. (2000). "Education for Growth: Why and For Whom?" NBER Working Paper #7591.

Masara, Sammy. (1996). *Kenya Education Directory 1996/97 Edition.* Nairobi: Express Communications.

Mauro, Paolo. (1995). "Corruption and Growth," *Quarterly Journal of Economics*, 110(3), 681-712.

McLeod P.L., S.A. Lobel SA, and T.H. Cox. (1996). "Ethnic diversity and creativity in small groups," *Small Group Research* 27 (2), 248-264.

Miguel, Edward, and Michael Kremer. (2000). "Child Health and Education: The Primary School Deworming Project in Kenya," mimeo., University of California at Berkeley.

Morgan, W.T.W., and N. Manfred Shaffer. (1966). *Population of Kenya, Density and Distribution: A geographical introduction to the Kenya Population Census, 1962*. Oxford University Press: Nairobi.

Ndegwa, Stephen. (1997). "Citizenship and Ethnicity: An Examination of Two Transition moments in Kenyan Politics," *American Political Science Review* 91(3), 599-616.

Platteau, Jean-Philippe. (1999). *Institutions, Social Norms, and Economic Development*. Harwood Press.

Putnam, Robert (with Robert Leonardi and Raffaela Nanetti). (1993). *Making Democracy Work*. Princeton University Press.

Simmel, Georg. (1971 [1908]). "Group Expansion and the development of individuality," in *Georg Simmel: On Individuality and Social Forms*, ed. Donald Levine (Chicago: University of Chicago Press).

Thomas, D.C. (1999). "Cultural diversity and work group effectiveness - An experimental study," *Journal of Cross-Cultural Psychology*, 30 (2) 242-263.

Were, Gideon. (1967). *A history of the Abaluyia of Western Kenya*. East African Publishing: Nairobi.

Wilson, L.S. (1992). "The *Harambee* movement and efficient public good provision in Kenya," *Journal of Public Economics*, 48(1), 1-19.

Woolcock, Michael. (1998). "Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework," *Theory and Society*, 27(20), 151-208.

## 7 Tables and Figures

Table 1: Selection into NGO Assistance program in 1995[31]

| | 1995 Pupil Enrollment (District Educational Office records) | | 1995 Average Government Exam Result, grades 6-8 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Explanatory variable | OLS | OLS | OLS | OLS |
| Indicator for selection into NGO assistance program | -86.5*** | -97.4** | -70.8*** | -48.3* |
| | (18.0) | (46.1) | (11.6) | (26.6) |
| Zonal residential ELF across tribes in 1996 | | 105.6 | | 135.5 |
| | | (114.4) | | (93.3) |
| (Indicator for selection into NGO assistance program) * | | 41.1 | | -108.6 |
| (Zonal residential ELF across tribes in 1996) | | (145.4) | | (84.7) |
| $R^2$ | 0.04 | 0.05 | 0.09 | 0.10 |
| Root MSE | 186.6 | 186.5 | 105.2 | 104.4 |
| Number of observations | 323 | 323 | 307 | 307 |
| Mean of dependent variable | 365.7 | | 871.3 | |
| (Standard deviation) | (190.6) | | (110.0) | |

Table 2: School Descriptive Statistics[32]

| | Mean | Standard deviation | Obs. |
|---|---|---|---|
| *School Characteristic* | | | |
| School ELF across tribes, 1996 Pupil Questionnaire data | 0.20 | 0.18 | 90 |
| School ELF across Luhya subtribes, 1996 Pupil Questionnaire data | 0.11 | 0.13 | 90 |
| Proportion largest ethnic group in school, 1996 Pupil Questionnaire data | 0.87 | 0.14 | 90 |
| School ELF across tribes, 1996 Exam Namelist data | 0.22 | 0.14 | 90 |
| ELF across tribes all schools within 5 km, 1996 Exam Namelist data | 0.24 | 0.14 | 90 |
| Total local school funds collected per pupil, 1995 (Kenyan Shillings) | 135.8 | 92.5 | 90 |
| *Harambee* donations collected per pupil, 1995 (Kenyan Shillings) | 45.1 | 84.4 | 90 |
| School fees collected per pupil, 1995 (Kenyan Shillings) | 90.7 | 43.5 | 90 |
| Official school fees per family, 1995 (Kenyan Shillings) | 310.5 | 144.6 | 90 |
| Desks per pupil, 1995 | 0.21 | 0.12 | 90 |
| Pupil latrines per pupil, 1995 | 0.016 | 0.013 | 90 |
| Classrooms per pupil, 1995 | 0.030 | 0.014 | 90 |
| School-owned texts per pupil, 1995 | 0.35 | 0.21 | 90 |
| Private texts per pupil, 1995 | 0.07 | 0.10 | 90 |
| Pupil enrollment per primary school, 1996 | 296.4 | 148.2 | 90 |
| Average score on 1996 NGO examination, grades 3-8 (in standard deviations) | 0.04 | 0.45 | 90 |
| Latitude (degrees north), GPS data | 0.43 | 0.19 | 90 |
| Longitude (degrees east), GPS data | 34.2 | 0.13 | 90 |
| Number of other primary schools within 5 km, GPS data | 13.5 | 3.8 | 90 |

[31] Data are from official District Education Office records. 100 of the 331 primary schools in Busia and Teso districts were selected for NGO assistance. The number of observations differs across regressions because not all schools have grade 6, 7, 8 classes; these schools having missing test scores.

[32] Data are from the 1996 ICS School and Pupil Questionnaires, 1996 Government Examination Namelists, and Global Positioning Systems (GPS) readings taken by NGO field workers. Ethno-linguistic fractionalization is defined as $1 - \sum_i$ (Proportion of Ethno-linguistic group$_i$ in the population)$^2$. School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group. School ELF across Luhya subtribes is defined as (ELF, Luhyas subtribes separate groups) – (ELF, Luhyas a single group).

Table 3: Geographic Zone Descriptive Statistics[33]

| | Mean | Standard deviation | Obs. |
|---|---|---|---|
| *Geographic Zone Characteristic* | | | |
| Zonal ELF across tribes, 1996 Pupil Questionnaire data | 0.23 | 0.14 | 22 |
| Zonal ELF across Luhya subtribes, 1996 Pupil Questionnaire data | 0.12 | 0.12 | 22 |
| Proportion largest ethnic group in zone, 1996 Pupil Questionnaire data | 0.86 | 0.11 | 22 |
| Proportion of Khayo (Luhya) pupils residing in the zone, 1996 | 0.17 | 0.24 | 22 |
| Proportion of Luo pupils residing in the zone, 1996 | 0.05 | 0.05 | 22 |
| Proportion of Marachi (Luhya) pupils residing in the zone, 1996 | 0.15 | 0.29 | 22 |
| Proportion of Nyala (Luhya) pupils residing in the zone, 1996 | 0.10 | 0.26 | 22 |
| Proportion of Samia (Luhya) pupils residing in the zone, 1996 | 0.19 | 0.33 | 22 |
| Proportion of Teso pupils residing in the zone, 1996 | 0.31 | 0.39 | 22 |
| Proportion of fathers with post-primary education in the zone, 1996 | 0.37 | 0.09 | 22 |
| Proportion of children in the zone with pit latrines at home, 1996 | 0.85 | 0.09 | 22 |
| Proportion of children in the zone with iron roofs at home, 1996 | 0.25 | 0.08 | 22 |
| Average number of full siblings among children in the zone, 1996 | 4.4 | 0.6 | 22 |
| Average years since school founding, 1996 | 22.6 | 5.2 | 22 |
| Average number of other primary schools within 5 km, GPS data | 13.0 | 3.5 | 22 |
| Average pupil enrollment per primary school, 1996 | 300.5 | 68.1 | 22 |
| Average pupil-teacher ratio in zone, 1996 | 28.9 | 4.7 | 22 |
| Proportion of teachers in zone with high school education, 1996 | 0.81 | 0.07 | 22 |
| Average years of teaching experience among teacher in zone, 1996 | 13.9 | 2.2 | 22 |
| Proportion of female teachers in the zone, 1996 | 0.26 | 0.13 | 22 |
| Proportion of headmaster in zone with high school education, 1996 | 0.90 | 0.17 | 22 |
| Average years at current school among headmaster in zone, 1996 | 5.1 | 2.7 | 22 |

[33] Data are from the 1996 ICS School, Teacher, and Pupil Questionnaires. Ethno-linguistic fractionalization is defined as $1 - \sum_i$ (Proportion of Ethno-linguistic group$_i$ in the population)$^2$. School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group. School ELF across Luhya subtribes is defined as (ELF, Luhyas subtribes separate groups) – (ELF, Luhyas a single group).

Table 4: Pupil Descriptive Statistics, by Ethnic Group[34]

| | Entire sample | Luhya pupils | Teso pupils | Luo pupils |
|---|---|---|---|---|
| Number of pupils interviewed | 6216 | 4114 | 1635 | 319 |
| Proportion of pupil sample | 1 | 0.67 | 0.26 | 0.05 |
| Age in years | 14.5 | 14.5 | 14.7 | 14.5 |
| Fathers with post-primary education | 0.38 | 0.38 | 0.38 | 0.33 |
| Fathers with formal employment | 0.31 | 0.33 | 0.23 | 0.41 |
| Proportion latrine ownership | 0.85 | 0.84 | 0.88 | 0.79 |
| Average number of full siblings | 4.5 | 4.4 | 4.6 | 4.1 |
| Proportion iron roof ownership | 0.25 | 0.26 | 0.22 | 0.30 |
| Attends primary school that is not the closest to home | 0.18 | 0.19 | 0.18 | 0.17 |
| Residence and school in different geographic zones | 0.12 | 0.15 | 0.06 | 0.07 |
| Lives with a parent, if at least one parent is alive | 0.85 | 0.84 | 0.88 | 0.83 |

Table 5: Ethnic Diversity across Geographic Divisions in Busia and Teso districts, in 1962 and 1996[35]

| Geographic division | *Name in 1962* | Proportion of largest residential ethnic group (Group in parentheses) | |
|---|---|---|---|
| | | *1962* | *1996 (Pupil Questionnaire data)* |
| Budalangi | *Bunyala* | 0.99 (Luhya) | 0.94 (Luhya) |
| Funyula | *Samia* | 0.98 (Luhya) | 0.94 (Luhya) |
| Butula | *Marachi* | 0.92 (Luhya) | 0.86 (Luhya) |
| Amukura/Chakol | *South Teso* | 0.92 (Teso) | 0.87 (Teso) |
| Angurai/Amagoro | *North Teso* | 0.87 (Teso) | 0.86 (Teso) |
| Nambale/Matayos | *Bukhayo* | 0.68 (Luhya) | 0.76 (Luhya) |

[34] Data are from the 1996 ICS Pupil Questionnaire administered to pupils in grades 6 to 8. Other ethnic groups, including Kalenjin, Kikuyu, Masaai, Somali, Tachoni, and Taita, comprise 0.020 of the sample.

[35] The 1962 data is from the 1962 Kenyan Census (Government of Kenya [1965]). These measures of ethnic diversity consider the Luhyas a single ethnic group, since the 1962 Census data does not contain information on Luhya subtribes. The 1996 data is from the ICS Pupil Questionnaire, which relies on self-described ethnic affiliation. Results using the 1996 Exam Namelist ethnicity information are similar.

Table 6: First Stage Regressions[36]

| | Dependent variable: | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | School ELF across tribes, Pupil Questionnaire data | | | | | | | School ELF across Luhya subtribes | 1– (Prop. largest group in school) |
| Explanatory variable | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS | (6) OLS | (7) OLS | (8) OLS | (9) OLS |
| Zonal residential ELF across tribes, Pupil Questionnaire data | 0.83*** (0.07) | 0.78*** (0.09) | 0.84*** (0.10) | 1.01*** (0.15) | | | | | |
| ELF across tribes for all schools within 5 km, Exam Namelist data | | | | | 0.85*** (0.10) | 0.88*** (0.10) | 1.07*** (0.14) | | |
| Zonal residential ELF across Luhya subtribes, Pupil Questionnaire data | | | | | | | | 0.72*** (0.18) | |
| 1– (Proportion largest ethnic group in zone), Pupil Questionnaire data | | | | | | | | | 0.85*** (0.06) |
| Geographic division indicators | No | Yes | Yes | Yes | No | No | Yes | Yes | Yes |
| Socioeconomic controls | No | No | Yes | Yes | No | Yes | Yes | Yes | Yes |
| Demographic and teacher controls | No | No | No | Yes | No | No | No | No | No |
| $R^2$ | 0.40 | 0.42 | 0.43 | 0.47 | 0.43 | 0.44 | 0.53 | 0.54 | 0.42 |
| Root MSE | 0.14 | 0.14 | 0.14 | 0.14 | 0.13 | 0.13 | 0.13 | 0.10 | 0.11 |
| Number of schools | 90 | 90 | 90 | 89 | 90 | 90 | 90 | 90 | 90 |
| Mean of dependent variable | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.11 | 0.87 |

[36] Huber robust standard errors in parentheses. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Regression disturbance terms are clustered at the zonal level. Ethno-linguistic fractionalization is defined as $1 - \sum_i$ (Proportion of Ethno-linguistic group$_i$ in the population)$^2$. School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group. School ELF across Luhya subtribes is defined as (ELF, Luhyas subtribes separate groups) – (ELF, Luhyas a single group). Socioeconomic controls include the proportion of fathers in the geographic zone with post-primary education, the proportion of pupils residing in the geographic zone with a latrine at home, and the average number of full siblings among pupils residing in geographic zone. Demographic controls include the proportions of Khayo (Luhya), Luo, Marachi (Luhya), Nyala (Luhya), Samia (Luhya), and Teso pupils residing in the geographic zone. Teacher controls include the proportion of teachers with a high school education, the average years of teaching experience, and the proportion of female teachers in the school.

Table 7: Reduced-form Regressions[37]

| Dependent variable | Coefficient estimate on zonal residential ELF across tribes, Questionnaire data (dependent variables are zonal averages) | Coefficient estimate on ELF across tribes among schools within 5 km, Exam Namelist data | Mean of dependent variable |
|---|---|---|---|
| | OLS | Spatial GMM | |
| Total local school funds collected per pupil in 1995 (Kenyan Shillings) | -185.5** | -150.0** | 135.8 |
| | (41.4) | (67.2) | |
| Proportion of fathers with post-primary education in geographic zone, 1996 | 0.07 | 0.08 | 0.38 |
| | (0.14) | (0.14) | |
| Proportion of fathers with formal sector employment in geographic zone, 1996 | -0.05 | -0.15*** | 0.31 |
| | (0.06) | (0.05) | |
| Average latrine ownership at home in the geographic zone, 1996 | 0.23** | 0.16** | 0.85 |
| | (0.11) | (0.08) | |
| Average iron roof ownership at home in the geographic zone, 1996 | 0.01 | -0.01 | 0.25 |
| | (0.09) | (0.10) | |
| Average number of full siblings for pupils in the geographic zone, 1996 | 0.5 | 1.0 | 4.5 |
| | (0.7) | (0.9) | |
| Pupil enrollment in school, geographic zone average (among sample schools) 1996 | 48.3 | -61.8 | 296.4 |
| | (95.3) | (102.2) | |
| Pupil-teacher ratio, geographic zone average 1996 | -6.4 | -12.1* | 28.9 |
| | (8.4) | (6.2) | |
| Proportion of teachers with HS education, zone average 1996 | 0.05 | -0.09 | 0.81 |
| | (0.09) | (0.13) | |
| Years of teaching experience, zone average 1996 | 0.8 | 3.3 | 13.9 |
| | (3.3) | (2.5) | |
| Proportion of female teachers, zone average 1996 | -0.01 | -0.06 | 0.26 |
| | (0.15) | (0.16) | |
| Proportion of headmasters with HS education, zone average 1996 | 0.21 | -0.21 | 0.90 |
| | (0.22) | (0.33) | |
| Years headmaster in current school, zone average 1996 | 0.1 | -3.9 | 5.1 |
| | (4.9) | (3.6) | |
| Number of observations | 22 geographic zones | 90 schools | |

[37] Huber robust standard errors in parentheses. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance in the spatial GMM specifications (Conley [1999]).

Table 8: Ethnic Diversity and Local School Funding
Dependent variable, Total local school funds collected per pupil in 1995 (Kenyan Shillings)[38]

| | Instrumental variables: Zonal residential ELF across tribes, 1996 Pupil Questionnaire data | | | | | | | Instrumental variable: ELF across tribes for schools within 5 km, 1996 Exam Namelist data | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Explanatory variable | (1) OLS | (2) IV-2SLS | (3) IV-2SLS | (4) IV-2SLS | (5) IV-2SLS | (6) IV-2SLS | (7) IV-2SLS | (8) Spatial GMM | (9) Spatial GMM | (10) Spatial GMM |
| School ELF across tribes | -22.7 (46.2) | -191.5** (95.1) | -103.6* (52.6) | -154.5*** (49.9) | -244.9* (127.1) | -191.3*** (50.3) | | -176.8** (86.1) | -95.8* (54.5) | -24.7 (54.1) |
| School ELF across Luhya subtribes | | | | | | -148.5 (107.3) | | | | |
| 1– (Proportion of largest ethnic group in school) | | | | | | | -158.3** (59.7) | | | |
| Proportion of fathers in the geographic zone with post-primary education | | | | 19.5 (86.4) | 197.8 (134.5) | 21.0 (83.9) | 16.5 (86.4) | | -104.6 (108.5) | -13.8 (114.9) |
| Proportion of pupils residing in the geographic zone with a latrine at home | | | | -174.2 (144.3) | -342.0** (121.8) | -291.5* (168.9) | -155.0 (138.3) | | -343.9 (226.7) | -119.4 (258.6) |
| Average number of full siblings among pupils residing in geographic zone | | | | -20.4 (16.5) | -7.7 (16.3) | -14.5 (18.2) | -19.9 (16.1) | | -7.6 (24.9) | -12.4 (28.7) |
| Geographic division indicators | No | No | Yes | Yes | Yes | Yes | Yes | No | No | Yes |
| Demographic and teacher controls | No | No | No | No | Yes | No | No | No | No | No |
| Root MSE | 92.9 | 97.6 | 86.8 | 90.1 | 97.0 | 90.4 | 88.3 | 96.8 | 90.1 | 85.7 |
| Number of schools | 90 | 90 | 90 | 90 | 89 | 90 | 90 | 90 | 90 | 90 |
| Mean dep. Variable | 135.8 | | | | | | | | | |

[38] Huber robust standard errors in parentheses. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones for Regressions 1 to 6. Ethno-linguistic fractionalization is defined as $1 - \sum_i$ (Proportion of Ethno-linguistic group$_i$ in the population)$^2$. School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group. School ELF across Luhya subtribes is defined as (ELF, Luhyas subtribes separate groups) – (ELF, Luhyas a single group). The instrumental variable for School ELF in Regressions 1 to 6 is ELF among pupils residing in the school's geographic zone. The instrumental variable for the proportion of the largest ethnic group in the school in Regression 7 is the proportion of the largest group among pupils residing in the geographic zone. The instrumental variable for School ELF in Regressions 8 to 10 is ELF among schools within 5 km of the school, using 1996 Exam Namelist data. Regression disturbance terms are allowed to be correlated across schools as a general function of physical distance in regressions 8 to 10 (Conley [1999]). Demographic controls include the proportions of Khayo (Luhya), Luo, Marachi (Luhya), Nyala (Luhya), Samia (Luhya), and Teso pupils residing in the geographic zone. Teacher controls include the proportion of teachers with a high school education, the average years of teaching experience, and the proportion of female teachers in the school.

Table 9: Additional Outcomes: School Funding, Facilities, Textbooks, and Test Scores[39]

| Dependent variable | Coefficient estimate on School ELF, IV-2SLS | Coefficient estimate on School ELF, Spatial GMM | Number of schools | Mean of dependent variable |
|---|---|---|---|---|
| *Local school funding* | | | | |
| Donations collected per pupil, 1995 (Kenyan Shillings) | -157.2* (76.1) | -116.1** (53.0) | 90 | 45.1 |
| School fees collected per pupil, 1995 (Kenyan Shillings) | 43.1 (36.1) | 20.3 (40.9) | 90 | 90.7 |
| *School facilities* | | | | |
| Desks per pupil, 1996 | -0.32*** (0.11) | -0.42*** (0.07) | 90 | 0.21 |
| Pupil latrines per pupil, 1996 | -0.009 (0.009) | -0.015 (0.011) | 90 | 0.016 |
| Classrooms per pupil, 1996 | -0.018 (0.019) | -0.018 (0.012) | 90 | 0.030 |
| Number of other primary schools within 5km | -13.5*** (4.7) | -15.1*** (3.0) | 90 | 13.5 |
| *Textbooks* | | | | |
| School-owned textbooks per pupil, 1996 | -0.21 (0.15) | -0.06 (0.16) | 90 | 0.35 |
| Privately-owned textbooks per pupil, 1996 | 0.01 (0.11) | 0.08 (0.08) | 90 | 0.07 |
| *Test scores* | | | | |
| Average school score on 1996 NGO exams, grades 3-8 (in standard deviations) | 0.03 (0.60) | 0.00 (0.46) | 90 | 0.04 |
| Geographic zone socioeconomic controls | Yes | Yes | | |

[39] Huber robust standard errors in parentheses. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Regression disturbance terms are clustered at the zonal level. Ethno-linguistic fractionalization is defined as $ELF \equiv 1 - \sum_i$ (Proportion of Ethno-linguistic group$_i$ in the population)$^2$. School ELF considers Luhyas a single group. The instrumental variable for School ELF in the IV-2SLS specifications is ELF among pupils residing in the school's geographic zone. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones for the IV-2SLS specifications. The instrumental variable for School ELF in the Spatial GMM specifications is ELF among schools within 5 km of the school using 1996 Exam Namelist data. Regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance in the Spatial GMM specifications, using the estimation strategy developed in Conley [1999]. Socioeconomic controls include the proportion of fathers in the geographic zone with post-primary education, the proportion of pupils residing in the geographic zone with a latrine at home, and the average number of full siblings among pupils residing in geographic zone.

Table 10: Ethnic diversity and Local School Funding within Geographic Zones[40]

| Dependent variable | Coefficient estimate on School ELF, OLS with Geographic Zone Fixed Effects | Number of schools | Mean of dependent variable |
|---|---|---|---|
| *Local school funding* | | | |
| Total local school funds collected per pupil, 1995 (Kenyan Shillings) | 89.7 (71.2) | 90 | 135.8 |
| Total local school funds collected per pupil, 1995 Geographic zones with higher than median ethnic diversity (zonal ELF >0.218) | 101.8 (69.9) | 45 | 107.6 |
| Total local school funds collected per pupil, 1995 Geographic zones with lower than median ethnic diversity (zonal ELF <0.218) | 63.6 (148.1) | 45 | 164.0 |
| School fees collected per pupil, 1995 (Kenyan Shillings) | 57.6* (34.4) | 90 | 90.7 |
| Donations collected per pupil, 1995 (Kenyan Shillings) | 32.1 (59.1) | 90 | 45.1 |
| *Test scores* | | | |
| Average school score on 1996 NGO exams, grades 3-8 (in standard deviations) | 0.72** (0.24) | 90 | 0.04 |
| Average school score on 1996 NGO exams, grades 3-8, Geographic zones with higher than median ethnic diversity (zonal ELF >0.218) | 1.01** (0.49) | 45 | 0.03 |
| Average school score on 1996 NGO exams, grades 3-8, Geographic zones with lower than median ethnic diversity (zonal ELF <0.218) | 0.09 (0.48) | 45 | 0.05 |
| *Pupil Population* | | | |
| Pupil enrollment in school, 1996 | 111.9 (126.1) | 90 | 296.4 |

[40] Huber robust standard errors in parentheses. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Regression disturbance terms are clustered at the zonal level. Ethno-linguistic fractionalization is defined as $ELF \equiv 1 - \sum_i$ (Proportion of Ethno-linguistic group$_i$ in the population)$^2$. School ELF across tribes considers Luhyas a single group.


Figure 1: Map of Kenya


Figure 2: Busia and Teso Districts, Kenya

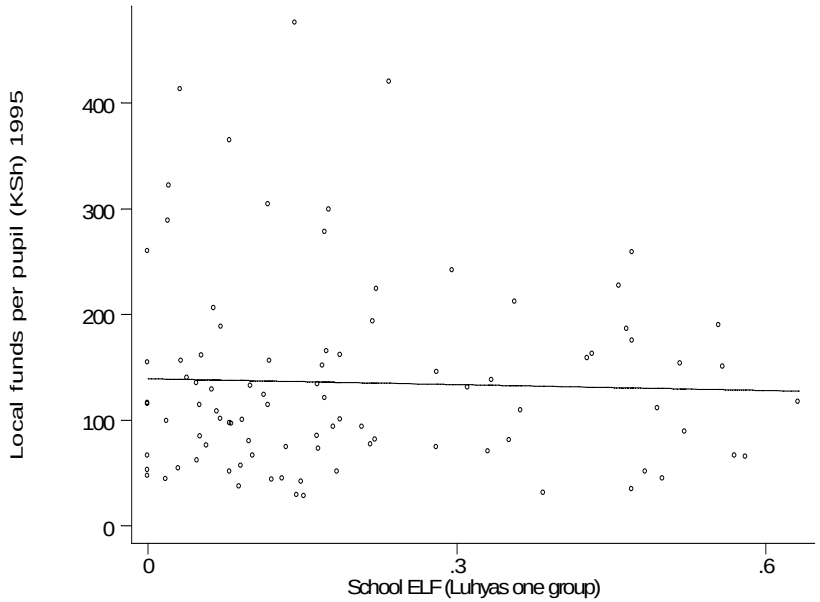


Figure 3: Total local funds collected per pupil in 1995 (Kenyan Shillings) versus school ethno-linguistic fractionalization (Pupil Questionnaire Data)[41]

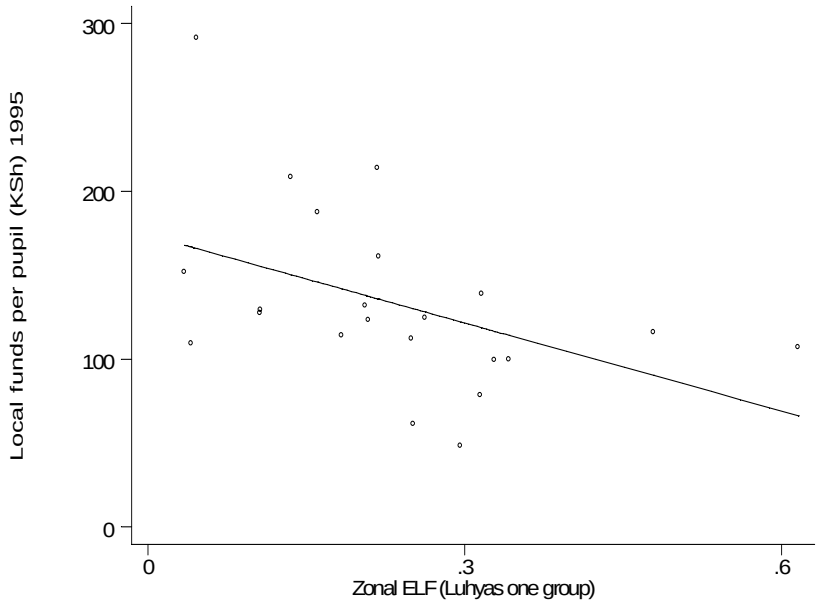


Figure 4: Total local school funds per pupil (Kenyan Shillings) in 1995 (geographic zone average) versus residential ethno-linguistic fractionalization in the geographic zone (Pupil Questionnaire Data)

[41] Figures 3 and 4 also contain the linear regression fits.

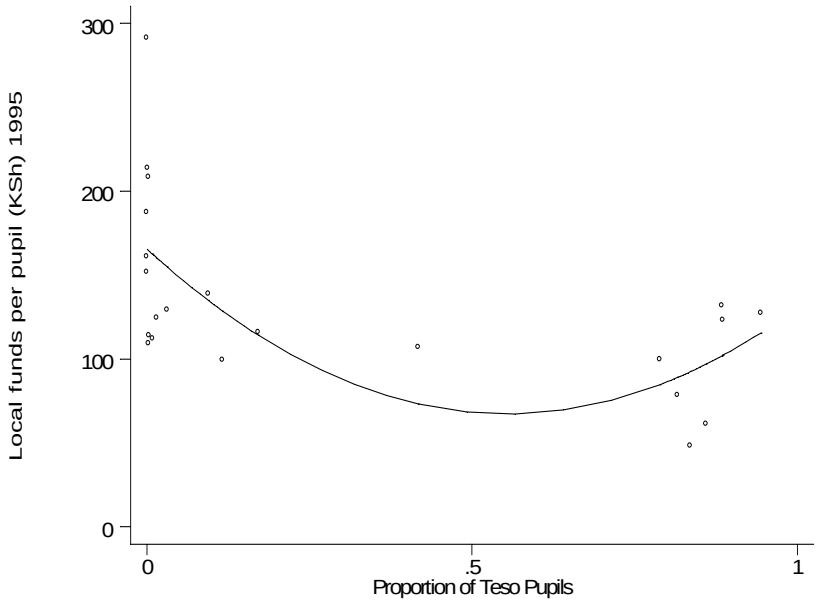


Figure 5: Total local school funds per pupil (Kenyan Shillings) in 1995 (geographic zone average) versus proportion of Teso pupils residing in the geographic zone (Pupil Questionnaire Data)[42]

[42] Figure5 also contains the quadratic regression fit.

## 8 Theory Appendix

**Proposition 1***:*
The mobility indifference condition (equation 5) is derived by equating $U^x(1) = U^x(2)$ and solving for $x^*$.

(a) High Diversity Units: $\underline{\phi} \le \phi \le 1$, where $\underline{\phi} = \frac{8}{9}$ when $d \in [-1/2, 1/2]$. In this case, *A* households are always the majority group in School 1 and *B* households are always the majority in School 2.

*(A1)* $$\gamma_1(x^*) = \frac{\int_0^{x^*} \phi x dx}{x^*} = \frac{\phi x^*}{2} \text{ and } \gamma_2(x^*) = \frac{\int_{x^*}^{1} (1-\phi x) dx}{1-x^*} = \left(1 - \frac{\phi}{2}\right) - \frac{\phi x^*}{2}$$

These imply that $x^*(e_1, e_2) = \frac{2 + (e_1 - e_2) - 2e_2 d\left(1 - \frac{\phi}{2}\right)}{4 - d\phi(e_1 + e_2)}$. A1 also leads to Equation A2:

*(A2)* $$\gamma_1(x^*) \ge \gamma_2(x^*) \Leftrightarrow x^* \ge \frac{\left(1 - \frac{\phi}{2}\right)}{\phi}$$

Inserting the expression for $x^*$ into Equation A2 implies that School 2 is (weakly) more ethnically diverse than School 1 for school quality realizations $(e_1, e_2) \in \{(0, 0), (1, 1), (0, 1)\}$, and that School 1 is more diverse for $(e_1, e_2) = (1, 0)$. The covariance of school funding and ethnic diversity is

*(A3)* $$Cov(g_s, \gamma_s) = E(g_s \cdot \gamma_s) - E(g_s) \cdot E(\gamma_s)$$

Equation A1 implies that $(\gamma_1 + \gamma_2)$ equals a constant for all $x^*$, which (slightly abusing notation) implies:

*(A4)*
$$\begin{aligned} Cov(g_s, \gamma_s) &= \frac{1}{8} \sum_{\forall (e_1, e_2)} [g_1(e_1, e_2) - g_2(e_1, e_2)][\gamma_1(e_1, e_2) - \gamma_2(e_1, e_2)] \\ &= \frac{1}{8}\left\{ d[\gamma_1(1,1) - \gamma_2(1,1)]^2 + g_1(1,0)[\gamma_1(1,0) - \gamma_2(1,0)] + g_2(0,1)[\gamma_1(0,1) - \gamma_2(0,1)] \right\} \end{aligned}$$

Each of the three terms in this expression is positive, yielding the result. □

*(b)* Low Diversity Units: $0 \le \phi \le \underline{\phi} < \underline{\phi}$ where $\underline{\phi} = \frac{8}{15}$ when $d \in [-1/2, 1/2]$. In this case, *A* households are always the majority group in School 1 and School 2. In this case:

*(A5)* $$\gamma_1(x^*) = \frac{\int_0^{x^*} \phi x dx}{x^*} = \frac{\phi x^*}{2} \text{ and } \gamma_2(x^*) = \frac{\int_{x^*}^{1} \phi x dx}{1-x^*} = \frac{\phi(1 + x^*)}{2}$$

These imply that $x^*(e_1, e_2) = \frac{2 + (e_1 - e_2) - e_2 d\phi}{4 - d\phi(e_1 - e_2)}$. A5 also implies Equation A6:

*(A6)* $$\gamma_1(x^*) \ge \gamma_2(x^*) \Leftrightarrow \frac{\phi}{2} \ge 0$$

Equation A6 implies that School 2 is more ethnically diverse than School 1 for all school quality realizations *($e_1$, $e_2$)*. In this case Equation A3 implies A7:

*(A7)* $$Cov(g_s, \gamma_s) = \frac{1}{8} \sum_{\forall(e_1,e_2)} \{g_1(e_1,e_2)[\gamma_1(e_1,e_2) - E(\gamma)] + g_2(e_1,e_2)[\gamma_2(e_1,e_2) - E(\gamma)]\}$$

Tedious but straightforward algebra yields an expression for the covariance; inserting possible values for *d* and $\phi$ indicates that the expression may be positive, negative, or zero.

**Proposition 2**:

*(b)* Proof: Inserting equation 5 into A1 for $\phi = 1$ implies that for high diversity units, $i, j \in \{1, 2\}$:

*(A8)* $$\gamma_{iu} = \frac{1}{2}\left(\frac{2 + (e_{iu} - e_{ju}) - d \cdot e_{ju}}{4 - d \cdot (e_{iu} + e_{ju})}\right) \text{ and } g_{iu} = \frac{1}{2} \cdot e_{iu} \cdot \left(\frac{4 + d \cdot (2 - d \cdot e_{ju})}{4 - d \cdot (e_{iu} + e_{ju})}\right)$$

Inserting equation 5 into A5 for $\phi = 0$ implies that for low diversity units,

*(A9)* $$\gamma_{iu} = 0 \text{ and } g_{iu} = \frac{1}{2} \cdot e_{iu}$$

The covariance of school funding and ethnic diversity across all schools is:

*(A10)* $$Cov(g_{su}, \gamma_{su}) = E(g_{su} \cdot \gamma_{su}) - E(g_{su}) \cdot E(\gamma_{su})$$

Inserting A8 and A9 into A10, and solving:

*(A11)* $$Cov(g_{su}, \gamma_{su}) = \frac{1}{64}\left(\frac{1}{(4-d)^2}\right) D\{(1-D)d(4-d)(10-d) + 4(d+2)^2\}$$

Setting A11 greater than zero and solving for *D* implies the inequality.

*(c)* Proof: Inserting equation 5 into A1 and taking averages implies that for high diversity units:

*(A12)* $$\overline{\gamma_u} = \frac{1}{2} \text{ and } \overline{g_u} = \frac{1}{2} \cdot \left(\frac{(2+d) \cdot (e_{iu} + e_{ju}) - d^2 \cdot e_{iu} \cdot e_{ju}}{4 - d \cdot (e_{iu} + e_{ju})}\right)$$

Inserting equation 5 into A5 and taking averages implies that for low diversity units,

*(A13)* $$\overline{\gamma_u} = 0 \text{ and } \overline{g_u} = \frac{1}{4} \cdot (e_{iu} + e_{ju})$$

The covariance of school funding and ethnic diversity across units is:

*(A14)* $$Cov(\overline{g_u}, \overline{\gamma_u}) = E(\overline{g_u} \cdot \overline{\gamma_u}) - E(\overline{g_u}) \cdot E(\overline{\gamma_u})$$

Inserting A12 and A13 into A14, and solving:

*(A15)* $$Cov(\overline{g_u}, \overline{\gamma_u}) = \frac{1}{64}\left(\frac{1}{(4-d)^2}\right) D\{(1-D)d(4-d)(10-d)\}$$

Since $d \in [-1/2, 1/2]$, it is straightforward to show that A15 is greater than zero when $d > 0$.

*(a)* Proof: Examination of A11 and A15 yields the result.