Table 1: Ethnic Diversity across Geographic Divisions in Busia and Teso districts, in 1962 and 1996\textsuperscript{41}

<table>
<thead>
<tr>
<th>Geographic division</th>
<th>Name in 1962</th>
<th>Proportion of largest residential ethnic group (Group in parentheses) 1962 (Pupil Questionnaire data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budalangi</td>
<td>Bunyala</td>
<td>0.99 (Luhya) 0.94 (Luhya)</td>
</tr>
<tr>
<td>Funyula</td>
<td>Samia</td>
<td>0.98 (Luhya) 0.94 (Luhya)</td>
</tr>
<tr>
<td>Butula</td>
<td>Marachi</td>
<td>0.92 (Luhya) 0.86 (Luhya)</td>
</tr>
<tr>
<td>Amukura/Chakol</td>
<td>South Teso</td>
<td>0.92 (Teso) 0.87 (Teso)</td>
</tr>
<tr>
<td>Angurai/Amagoro</td>
<td>North Teso</td>
<td>0.87 (Teso) 0.86 (Teso)</td>
</tr>
<tr>
<td>Nambale/Matayos</td>
<td>Bukhayo</td>
<td>0.68 (Luhya) 0.76 (Luhya)</td>
</tr>
</tbody>
</table>

Table 2: Pupil Descriptive Statistics, by Ethnic Group\textsuperscript{42}

<table>
<thead>
<tr>
<th></th>
<th>Entire sample</th>
<th>Luhya pupils</th>
<th>Teso pupils</th>
<th>Luo pupils</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pupils interviewed</td>
<td>5832</td>
<td>3867</td>
<td>1516</td>
<td>301</td>
</tr>
<tr>
<td>Proportion of pupil sample</td>
<td>1</td>
<td>0.66</td>
<td>0.26</td>
<td>0.05</td>
</tr>
<tr>
<td>Age in years</td>
<td>14.5</td>
<td>14.4</td>
<td>14.7</td>
<td>14.5</td>
</tr>
<tr>
<td>Father years of education</td>
<td>7.5</td>
<td>7.5</td>
<td>7.4</td>
<td>7.5</td>
</tr>
<tr>
<td>Mother years of education</td>
<td>5.0</td>
<td>5.0</td>
<td>4.9</td>
<td>5.2</td>
</tr>
<tr>
<td>Fathers with formal employment\textsuperscript{43}</td>
<td>0.24</td>
<td>0.26</td>
<td>0.17</td>
<td>0.29</td>
</tr>
<tr>
<td>Mothers with formal employment\textsuperscript{40}</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Proportion latrine ownership</td>
<td>0.85</td>
<td>0.84</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>Proportion iron roof ownership</td>
<td>0.26</td>
<td>0.26</td>
<td>0.23</td>
<td>0.30</td>
</tr>
<tr>
<td>Proportion livestock ownership\textsuperscript{44}</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.80</td>
</tr>
<tr>
<td>Proportion cultivates corn (maize)</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.92</td>
</tr>
<tr>
<td>Proportion cultivates cash crop\textsuperscript{45}</td>
<td>0.39</td>
<td>0.33</td>
<td>0.56</td>
<td>0.29</td>
</tr>
<tr>
<td>Attends primary school that is not the closest to home</td>
<td>0.18</td>
<td>0.19</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>Residence and school in different geographic zones</td>
<td>0.12</td>
<td>0.15</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Lives with a parent, if at least one parent is alive</td>
<td>0.85</td>
<td>0.84</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>Average number of full siblings</td>
<td>4.5</td>
<td>4.4</td>
<td>4.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Proportion Catholic</td>
<td>0.56</td>
<td>0.57</td>
<td>0.56</td>
<td>0.64</td>
</tr>
</tbody>
</table>

\textsuperscript{41} The 1962 data is from the 1962 Kenyan Census (Government of Kenya 1965). The 1996 data is from the ICS Pupil Questionnaire, which relies on self-described ethnic affiliation.

\textsuperscript{42} Data are from the 1996 ICS Pupil Questionnaire administered to pupils in grades 6 to 8. Other ethnic groups, including Kalenjin, Kikuyu, Masai, Somali, Tachoni, and Taita, comprise 0.020 of the sample.

\textsuperscript{43} Formal employment includes government, factory, and corporate jobs, but not petty trading (in contrast to the definition of formal employment in Miguel 2001).

\textsuperscript{44} Livestock ownership includes cattle, goats, sheep, and pigs (but not chicken).

\textsuperscript{45} Cash crops include tobacco, sugar cane, and cotton.
Table 5: Ethnic Diversity and Local Primary School Funding

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>School ELF across tribes</th>
<th>Dependent variable</th>
<th>Total local primary school funds collected per pupil in 1995 (Kenyan Shillings)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st stage</td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td>Ethnic diversity measures</td>
<td>School ELF across tribes</td>
<td>-32.9</td>
<td>-216.4**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(64.0)</td>
<td>(88.4)</td>
</tr>
<tr>
<td></td>
<td>Zonal ELF across tribes</td>
<td>0.86***</td>
<td>-185.7**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(77.9)</td>
</tr>
<tr>
<td></td>
<td>1 – (Proportion largest ethnic group in zone)</td>
<td></td>
<td>-162.9**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(66.6)</td>
<td>(76.3)</td>
</tr>
<tr>
<td></td>
<td>ELF across tribes for all schools within 5 km</td>
<td></td>
<td>-174.0**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(76.3)</td>
<td>(80.8)</td>
</tr>
<tr>
<td>Zonal controls</td>
<td>Proportion fathers with formal employment</td>
<td>189.5</td>
<td>-220.6*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(165.1)</td>
<td>(120.5)</td>
</tr>
<tr>
<td></td>
<td>Proportion of pupils with a latrine at home</td>
<td>-431.6***</td>
<td>-286.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(139.9)</td>
<td>(228.0)</td>
</tr>
<tr>
<td></td>
<td>Proportion livestock ownership</td>
<td>120.1</td>
<td>186.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(136.9)</td>
<td>(130.4)</td>
</tr>
<tr>
<td></td>
<td>Proportion cultivates cash crop</td>
<td>35.7</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(61.4)</td>
<td>(106.9)</td>
</tr>
<tr>
<td></td>
<td>Proportion Teso pupils</td>
<td>67.9</td>
<td>(181.4)</td>
</tr>
<tr>
<td>Geographic division indicators</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Root MSE</td>
<td>0.14</td>
<td>99.8</td>
<td>96.7</td>
</tr>
<tr>
<td>R²</td>
<td>0.40</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Number of schools</td>
<td>84</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>Mean dep. variable</td>
<td>0.20</td>
<td>152.6</td>
<td>152.6</td>
</tr>
</tbody>
</table>

50 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 7. Ethno-linguistic fractionalization is defined as \( I - \sum(\text{Proportion of Ethno-linguistic group i in the population})^2 \). School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhya as a single group. Regression disturbance terms are allowed to be correlated across schools as a general function of physical distance in regressions 8 and 9 (Conley 1999). Geographic indicators are indicators for six (of the seven) geographic divisions.
<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Coefficient estimate on zonal residential ELF across tribes (OLS)</th>
<th>Coefficient estimate on ELF across tribes among schools within 5 km (Spatial OLS)</th>
<th>Number of schools</th>
<th>Mean dependent variable</th>
</tr>
</thead>
</table>
| **Local school funding** | -157.1**  
1995 (Kenyan Shillings)  
(61.6) | -182.1**  
(68.5) | 84 | 44.8 |
| School fees collected per pupil, 1995  
(Kenyan Shillings) | 11.9  
(35.2) | 8.1  
(64.6) | 84 | 107.8 |
| **School facilities, inputs** | | | | |
| Desks per pupil, 1996 | -0.20**  
(0.08) | -0.31***  
(0.08) | 84 | 0.21 |
| Pupil latrines per pupil, 1996 | -0.007  
(0.009) | -0.007  
(0.013) | 84 | 0.016 |
| Classrooms per pupil, 1996 | -0.016  
(0.016) | -0.023  
(0.013) | 84 | 0.030 |
| School-owned textbooks per pupil, 1996 | -0.17  
(0.13) | -0.27  
(0.17) | 84 | 0.34 |
| Private texts (at home) per pupil, 1996 | -0.03  
(0.07) | -0.10  
(0.09) | 84 | 0.07 |
| **Number of other primary schools within 5km** | -10.2***  
(3.5) | -12.2***  
(3.7) | 84 | 14.5 |
| **Test scores** | | | | |
| Average school score on 1996 NGO exams, grades 3-8 (in standard deviations) | 0.10  
(0.52) | 0.11  
(0.52) | 84 | 0.05 |
| **Socioeconomic controls (zonal averages)** | Yes | Yes | | |

---

51 Huber robust standard errors in parentheses. Significantly different than zero at 90% (*), 95% (**), 99% (***). Regression disturbance terms are clustered at the zonal level. Ethno-linguistic fractionalization is defined as $ELF = 1 - \Sigma_i (Proportion of Ethno-linguistic group i in the population)^2$. School ELF considers Luhyas a single group.

52 1996 Pupil Questionnaire Data. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones.

53 1996 Exam Name list data. Regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance, using the estimation strategy developed in Conley (1999).

54 Socioeconomic controls include the proportion of fathers in the geographic zone with formal sector employment, the proportion of pupils residing in the geographic zone with a latrine at home, the proportion of pupils whose households own livestock, and the proportion of pupils whose households cultivate a cash crop. The test score results also an additional explanatory variable, an indicator for having received financial assistance through another NGO program.
Table 8: Well Descriptive Statistics\textsuperscript{59}

<table>
<thead>
<tr>
<th>Well Characteristic</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELF across tribes for all primary schools within 5 km of the well, 1996 Exam Namelist data</td>
<td>0.23</td>
<td>0.14</td>
<td>667</td>
</tr>
<tr>
<td>Indicator variable “normal” water flow from well, 2000-2001 survey</td>
<td>0.57</td>
<td>0.49</td>
<td>667</td>
</tr>
<tr>
<td>Indicator variable no broken or missing well parts, 2000-2001 survey\textsuperscript{60}</td>
<td>0.66</td>
<td>0.48</td>
<td>667</td>
</tr>
<tr>
<td>Indicator variable people in the area get water from another local well (if not normal water flow), 2000-2001 survey</td>
<td>0.32</td>
<td>0.47</td>
<td>196</td>
</tr>
<tr>
<td>Year well stopped functioning (if not normal water flow), 2000-2001 survey</td>
<td>1997.5</td>
<td>3.1</td>
<td>196</td>
</tr>
<tr>
<td>Latitude (degrees North), GPS data from 2000-2001 survey</td>
<td>0.36</td>
<td>0.17</td>
<td>667</td>
</tr>
<tr>
<td>Longitude (degrees East), GPS data from 2000-2001 survey</td>
<td>34.20</td>
<td>0.12</td>
<td>667</td>
</tr>
</tbody>
</table>

Table 9: Ethnic Diversity and Well Maintenance\textsuperscript{61}

<table>
<thead>
<tr>
<th>Indicator variable for “normal” water flow from well</th>
<th>Dependent variable:</th>
<th>Indicator variable for no broken or missing well parts</th>
<th></th>
<th>Indicator variable people in the area get water from another local well (if the KEFINCO well does not have normal water flow)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Probit</td>
<td>(2) Spatial OLS</td>
<td>(3) Probit</td>
<td>(4) Spatial OLS</td>
<td>(5) Probit</td>
</tr>
<tr>
<td>ELF across tribes among schools within 5 km</td>
<td>-0.26\textsuperscript{*}</td>
<td>-0.26</td>
<td>-0.25\textsuperscript{*}</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.17)</td>
<td>(0.13)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Number of wells</td>
<td>667</td>
<td>667</td>
<td>667</td>
<td>667</td>
</tr>
<tr>
<td>Root MSE</td>
<td>-</td>
<td>0.49</td>
<td>-</td>
<td>0.47</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.57</td>
<td>0.57</td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

\textsuperscript{59} 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 $I - \sum (\text{Proportion of Ethno-linguistic group, in the population})^2$. School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group.

\textsuperscript{60} Well parts include the pump handle, the cover and base, and the external and internal pipes and seals.

\textsuperscript{61} Notes: Huber robust standard errors in parentheses. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones in regressions 1, 3, and 5, and regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance, using the estimation strategy developed in Conley (1999), in regressions 2, 4, and 6. Significant at 90\textsuperscript{(*)}, 95\textsuperscript{(**)}, 99\textsuperscript{(***)} % confidence. Geographic indicators are indicators for six (of the seven) geographic divisions.
Table 7: School committee records and field officer observations

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Coefficient estimate on zonal residential ELF across tribes (OLS)</th>
<th>Coefficient estimate on ELF across tribes among schools within 5 km (Spatial OLS)</th>
<th>Number of schools</th>
<th>Mean dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Committee Records</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School committee record items regarding sanctions or verbal pressure, 1997</td>
<td>-3.7**</td>
<td>-4.2*</td>
<td>84</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>(1.6)</td>
<td>(2.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School committee record items regarding administrative activities, 1997</td>
<td>5.7</td>
<td>6.2</td>
<td>84</td>
<td>18.9</td>
</tr>
<tr>
<td></td>
<td>(6.1)</td>
<td>(10.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent school meetings, 1997</td>
<td>-1.6</td>
<td>-1.3</td>
<td>84</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>(1.1)</td>
<td>(1.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field Officer Observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent cooperation from 0 to 1 (reported by field officers), 1998</td>
<td>-0.77**</td>
<td>-0.84**</td>
<td>84</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher motivation from 0 to 1 (reported by field officers), 1998</td>
<td>-0.39*</td>
<td>-0.49*</td>
<td>84</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic controls (zonal averages)</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

55 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 $ELF = 1 - \sum (Proportion of Ethno-linguistic group, in the population)^2$. School ELF considers Luhyas a single group.

56 1996 Pupil Questionnaire Data. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones.

57 1996 Exam Namelist data. Regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance, using the estimation strategy developed in Conley (1999).

58 Socioeconomic controls include the proportion of fathers in the geographic zone with formal sector employment, the proportion of pupils residing in the geographic zone with a latrine at home, the proportion of pupils whose households own livestock, and the proportion of pupils whose households cultivate a cash crop.
### Table 3: Group Summary Statistics

Means and Standard Deviations

<table>
<thead>
<tr>
<th>Method of Arrival to Group</th>
<th>Uninvited to Group</th>
<th>Invited to Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean &amp; Std Error</td>
<td>Mean &amp; Std Error</td>
</tr>
<tr>
<td></td>
<td># of Obs</td>
<td># of Obs</td>
</tr>
</tbody>
</table>

#### GEOGRAPHIC CONCENTRATION

- *Average distance to original members from current members (minutes)*
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12.858</td>
<td>0.785</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>5.088</td>
<td>0.694</td>
<td>42</td>
</tr>
</tbody>
</table>

- *Average percent of original members who live within 10 minutes of current member*
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.228</td>
<td>0.028</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.182</td>
<td>0.027</td>
<td>42</td>
</tr>
</tbody>
</table>

- GD: Geographic concentration
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.147</td>
<td>0.016</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.104</td>
<td>0.027</td>
<td>42</td>
</tr>
</tbody>
</table>

- E(GD): Expected geographic concentration
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.127</td>
<td>0.014</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.090</td>
<td>0.026</td>
<td>42</td>
</tr>
</tbody>
</table>

#### CULTURAL CONCENTRATION

- *Average percent of original members of same culture as current member*
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.200</td>
<td>0.014</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.092</td>
<td>0.013</td>
<td>42</td>
</tr>
</tbody>
</table>

- CD: Cultural concentration
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.119</td>
<td>0.021</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.136</td>
<td>0.024</td>
<td>42</td>
</tr>
</tbody>
</table>

- E(CD): Expected cultural concentration
  
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Error</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.106</td>
<td>0.012</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.078</td>
<td>0.021</td>
<td>42</td>
</tr>
</tbody>
</table>

---

*Variables with asterisks are the key dependent variables used in the specifications in Tables 3, 4, and 5.
All results calculated on original group members only.

\[
GD_{group} = \sum_{neighborhoods} (s_i - \bar{x})^2,
\]

where \( s_i \) is the share of the group from neighborhood \( i \) and \( \bar{x} \) is the share of the general population from neighborhood \( i \).

\[
E(GD)_{group} = [1 - \sum_{neighborhoods} (x_i)^2] / n
\]

CD and E(CD) are constructed identically to GD and E(GD), except by cultural group rather than neighborhood.

The Alesina index for cultural concentration is equal to the sum of squared shares of each cultural group.
Table 4: Individual Default
OLS, Tobit, and Probit

<table>
<thead>
<tr>
<th>Dependent variable: Percent of loan in default at end of cycle</th>
<th>1st Loan Only</th>
<th>All Loans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Tobit</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Distance from individual's home to original members of group</td>
<td>0.019</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.342)</td>
</tr>
<tr>
<td></td>
<td>n=616</td>
<td>n=616</td>
</tr>
<tr>
<td>Percent of original members within 10-minute walk of individual's home</td>
<td>-1.536 ***</td>
<td>-6.077 ***</td>
</tr>
<tr>
<td></td>
<td>(0.391)</td>
<td>(1.795)</td>
</tr>
<tr>
<td></td>
<td>n=616</td>
<td>n=616</td>
</tr>
<tr>
<td>Percent of original members with same culture as individual</td>
<td>-0.534 *</td>
<td>-4.230 **</td>
</tr>
<tr>
<td></td>
<td>(0.301)</td>
<td>(1.791)</td>
</tr>
<tr>
<td></td>
<td>n=616</td>
<td>n=616</td>
</tr>
</tbody>
</table>

*** 99% significance; ** 95% significance; * 90% significance
Each cell is a separate specification.
Standard errors corrected for clustering at the group level in all specifications.
Individuals weighted evenly "all loans" specifications.
Individual level specifications include the following control variables (See Appendix Table 2 for results on control variables):
Distance to FINCA (town center), town dummy, neighborhood dummies, age, education, marital status, siblings, children,
# in household, year, and age of group when individual joined.
Loan size estimated using approved loan amount, which is savings balance at end of prior cycle.
Table 6: Dropout Probit

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0.115***</td>
<td>0.112***</td>
<td>0.113***</td>
<td>-0.023</td>
<td>0.153***</td>
<td>0.197***</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.036)</td>
<td>(0.056)</td>
<td>(0.043)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Total Accumulated Savings</td>
<td>-0.013</td>
<td>-0.016</td>
<td>-0.014</td>
<td>-0.014</td>
<td>-0.015</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Distance from individual's home to original members of group</td>
<td>0.037</td>
<td>0.023</td>
<td>(0.031)</td>
<td>(0.032)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance Interacted with default</td>
<td>0.074***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of original members within 10-minute walk of individual's home</td>
<td>-0.007</td>
<td></td>
<td></td>
<td>-0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent within 5-minute walk Interacted with default</td>
<td></td>
<td>-0.132***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of original members with same culture as individual</td>
<td></td>
<td></td>
<td>-0.192</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.157)</td>
<td></td>
<td>(0.144)</td>
<td></td>
</tr>
<tr>
<td>Culture Interacted with default</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.332***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.094)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>616</td>
<td>616</td>
<td>616</td>
<td>616</td>
<td>616</td>
<td>616</td>
</tr>
<tr>
<td># of dropouts</td>
<td>148</td>
<td>148</td>
<td>148</td>
<td>148</td>
<td>148</td>
<td>148</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-173.47</td>
<td>-173.76</td>
<td>-173.28</td>
<td>-167.39</td>
<td>-171.58</td>
<td>-166.78</td>
</tr>
<tr>
<td>Groups</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>

*** 99% significance; ** 95% significance; * 90% significance
Marginal effects of probit reported.
Standard errors corrected for clustering at the group level.
Individual-level specifications control variables for distance to FINCA (town center), town dummy, neighborhood dummies, age, education, marital status, siblings, children, # in household, year, and age of group.
Figure 3: Map of Baltistan with sampled communities indicated

Figure 4: Graphical Illustration of Claim 1

Marginal $S$

Marginal Cost

Marginal Benefit $^{A_2}$

Marginal Benefit $^{A_1}$

Marginal Benefit $^B$

Aggregate Household Labor = $l_i + m_i$
Table 5. Determinants of Maintenance

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) OLS</th>
<th>Variables</th>
<th>(1-cont) OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Community factors:</strong></td>
<td></td>
<td><strong>Physical capital variables:</strong></td>
<td></td>
</tr>
<tr>
<td>Land Inequality</td>
<td>-275.4**</td>
<td>Mean off-farm hh Income</td>
<td>-.0012</td>
</tr>
<tr>
<td>(117)</td>
<td></td>
<td>(0.0021)</td>
<td></td>
</tr>
<tr>
<td>Land Inequality Squared</td>
<td>395.7**</td>
<td>Mean real estate value (000)</td>
<td>-.044**</td>
</tr>
<tr>
<td>(194)</td>
<td></td>
<td>(0.022)</td>
<td></td>
</tr>
<tr>
<td>Social Heterogeneity</td>
<td>-55.1***</td>
<td>Community wage</td>
<td>.127</td>
</tr>
<tr>
<td>(18)</td>
<td></td>
<td>(0.160)</td>
<td></td>
</tr>
<tr>
<td>Community size</td>
<td>-.020</td>
<td>No. of Mechanical asset households</td>
<td>2.05***</td>
</tr>
<tr>
<td>(0.62)</td>
<td></td>
<td>(0.49)</td>
<td></td>
</tr>
<tr>
<td>Total Cultivable land</td>
<td>2e-04</td>
<td>Access to Electricity?</td>
<td>5.13</td>
</tr>
<tr>
<td>(18e-04)</td>
<td></td>
<td>(4.38)</td>
<td></td>
</tr>
<tr>
<td>Single cropping zone?</td>
<td>-8.79</td>
<td>Access to Health facility?</td>
<td>-1.21</td>
</tr>
<tr>
<td>(5.58)</td>
<td></td>
<td>(4.50)</td>
<td></td>
</tr>
<tr>
<td>Walk Time</td>
<td>-.046</td>
<td>Access to Potable Water?</td>
<td>3.22</td>
</tr>
<tr>
<td>(0.075)</td>
<td></td>
<td>(4.37)</td>
<td></td>
</tr>
<tr>
<td>Travel Time</td>
<td>-.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Human capital variables:</strong></td>
<td></td>
<td><strong>Project factors:</strong></td>
<td></td>
</tr>
<tr>
<td>Shopkeepers fraction</td>
<td>-71.9***</td>
<td>Project New?</td>
<td>-20.5***</td>
</tr>
<tr>
<td>(27.2)</td>
<td></td>
<td>(6.05)</td>
<td></td>
</tr>
<tr>
<td>Skilled workers fraction</td>
<td>32.5**</td>
<td>Government project?</td>
<td>-8.96</td>
</tr>
<tr>
<td>(15.0)</td>
<td></td>
<td>(7.58)</td>
<td></td>
</tr>
<tr>
<td>Basic Education fraction</td>
<td>-3.37</td>
<td>Project Leader exists?</td>
<td>11.3**</td>
</tr>
<tr>
<td>(3.72)</td>
<td></td>
<td>(5.7)</td>
<td></td>
</tr>
<tr>
<td>Tertiary Education fraction</td>
<td>36.0</td>
<td>External Funds (000,000)</td>
<td>6.96</td>
</tr>
<tr>
<td>(26.3)</td>
<td></td>
<td>(11.3)</td>
<td></td>
</tr>
<tr>
<td>Religious Education fraction</td>
<td>-36.8</td>
<td>Project Complexity</td>
<td>-4.20</td>
</tr>
<tr>
<td>(24.1)</td>
<td></td>
<td>(2.77)</td>
<td></td>
</tr>
<tr>
<td>High school?</td>
<td>20.6***</td>
<td>Non-technical decisions participation</td>
<td>35.5***</td>
</tr>
<tr>
<td>(6.4)</td>
<td></td>
<td>(15.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Technical decisions participation</td>
<td>-18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Controls</td>
<td>Pj age, type</td>
</tr>
<tr>
<td>Adj R²</td>
<td>.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prob&gt;F</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>132</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Huber-White robust standard errors in parentheses
Disturbance terms clustered at the village level
***Significantly different from zero at 1%
**Significantly different from zero at 5%
* Significantly different from zero at 10%

---

*Project Share Inequality* was not included in these regressions since it is highly correlated with *Land Inequality*. 38
The residuals are calculated from a regression of project maintenance on project age, type and community dummy variables only. Excluding the two outlyers (highest inequality levels) does change the shape or fit of the plot.

Predicted effects are calculated by estimating the marginal effect of the inequality measure after controlling for all other factors. For the land inequality the estimates used are from the regression in Table 5, and for project share inequality the estimates are from the regression in Column 1, Table 7. Excluding the high project share inequality outlyer does not significantly change the point estimates in the regression.
Table 6. Effect of Project Leadership on Maintenance
OLS and Instrumental Variable (IV-2SLS)

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) 1st stage OLS</th>
<th>(2) 2nd stage IV-2SLS</th>
<th>(3) OLS</th>
<th>(4) 1st stage OLS</th>
<th>(5) 2nd stage IV-2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Leader exists?</td>
<td>Dependent variable</td>
<td>32.64**</td>
<td>36.45***</td>
<td></td>
<td>25.11*</td>
</tr>
<tr>
<td>Leader Quality</td>
<td></td>
<td>(15.70)</td>
<td>(13.64)</td>
<td></td>
<td>(15.04)</td>
</tr>
<tr>
<td>Hereditary family 25-50 healthy male?</td>
<td>0.30*</td>
<td>(0.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hereditary family absence (1-3)</td>
<td>-.32***</td>
<td>(.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hereditary family average age</td>
<td>-.015**</td>
<td>(.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hereditary family 25-50 educated, present male</td>
<td></td>
<td></td>
<td></td>
<td>0.10***</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Hereditary family non-farm?</td>
<td></td>
<td></td>
<td></td>
<td>-0.20***</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Ideal leaders in community? (1-4)</td>
<td></td>
<td></td>
<td></td>
<td>0.21***</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Project Leader attributes:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educated?</td>
<td></td>
<td>-3.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>-.52**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-farm occupation?</td>
<td>-10.4**</td>
<td>(5.95)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land holding</td>
<td>-.001</td>
<td>(.056)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present throughout year?</td>
<td>7.63</td>
<td>(9.22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trained?</td>
<td></td>
<td>5.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From hereditary leader group?</td>
<td></td>
<td>-10.4*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj R²</td>
<td>.09</td>
<td>.21</td>
<td>.34</td>
<td>.07</td>
<td>.43</td>
</tr>
<tr>
<td>Prob&gt;F</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>N</td>
<td>132</td>
<td>132</td>
<td>130</td>
<td>132</td>
<td>132</td>
</tr>
</tbody>
</table>

Huber-White robust standard errors in parentheses
Disturbance terms clustered at the village level
***Significantly different from zero at 1%
**Significantly different from zero at 5%
* Significantly different from zero at 10%
### Table 7. Project-specific Determinants of Maintenance

**Community Fixed Effects**

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) FE</th>
<th>(2) FE</th>
<th>(3) FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Complexity</td>
<td>-12.76***</td>
<td>-15.19***</td>
<td>-15.44***</td>
</tr>
<tr>
<td></td>
<td>(3.85)</td>
<td>(3.04)</td>
<td>(3.92)</td>
</tr>
<tr>
<td>Project Share Inequality</td>
<td>-373.3***</td>
<td>-402.7***</td>
<td>-422***</td>
</tr>
<tr>
<td></td>
<td>(67.7)</td>
<td>(86)</td>
<td>(69.2)</td>
</tr>
<tr>
<td>Project Share Inequality squared</td>
<td>1304****</td>
<td>1391***</td>
<td>1381***</td>
</tr>
<tr>
<td></td>
<td>(225)</td>
<td>(267)</td>
<td>(211)</td>
</tr>
<tr>
<td>Non-technical decisions participation</td>
<td>55.43*</td>
<td>50.87*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(28.29)</td>
<td>(24.24)</td>
<td></td>
</tr>
<tr>
<td>Technical decisions participation</td>
<td>-38.49*</td>
<td>-34*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.56)</td>
<td>(16.68)</td>
<td></td>
</tr>
<tr>
<td>Government project?</td>
<td>-23.63***</td>
<td>-18***</td>
<td>-18.18**</td>
</tr>
<tr>
<td></td>
<td>(7.95)</td>
<td>(6.07)</td>
<td>(8.03)</td>
</tr>
<tr>
<td>Project New?</td>
<td>-41.92***</td>
<td>-40.55***</td>
<td>-46.77***</td>
</tr>
<tr>
<td></td>
<td>(13.67)</td>
<td>(11)</td>
<td>(15.06)</td>
</tr>
<tr>
<td>Project Leader?</td>
<td></td>
<td></td>
<td>13.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(8.42)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controls</th>
<th>Community Fixed Effects, Project Age and type</th>
<th>Community Fixed Effects, Project Age and type</th>
<th>Community Fixed Effects, Project Age and type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj R²</td>
<td>.71</td>
<td>.63</td>
<td>.73</td>
</tr>
<tr>
<td>Prob&gt;F</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>N</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
</tbody>
</table>

Column 1 presents the primary regression. Column 2 checks to see whether the results remain similar once the potentially endogenous (Halo effects) participation measure is excluded. Column 3 checks to see whether the external agency effect remains once leadership presence (endogenous) and participation are both controlled for.

Huber-White robust standard errors in parentheses

***Significantly different from zero at 1%
**Significantly different from zero at 5%
* Significantly different from zero at 10%