The Effect of Proximity on School Enrollment: Evidence from a RCT in Afghanistan[1]

Preliminary

Please Do Not Circulate

March 2009

Dana Burde
(NYU Steinhardt)

Leigh L. Linden
(Columbia, IZA)

Abstract:
We conduct a randomized evaluation to assess the causal effect of distance on children's academic participation and performance. Using a sample of 31 villages and over 1,500 children in rural northwestern Afghanistan, we randomly assign 13 villages to receive a community-based school. The program significantly increases enrollment and test scores amongst all children and dramatically improves the existing gender disparities. The intervention increases formal school enrollment by 47 percentage points among all children and increases their average test scores by 0.59 standard deviations. For those children actually participating in the schools, average test scores increase by 1.2 standard deviations. Overall, children prove very sensitive to changes in the distance to the nearest school. Enrollment falls by 16 percentage points for every mile that children must travel to school and test scores fall by 0.19 standard deviations. Girls benefit more from the program than boys, increasing formal school enrollment by 15 percentage points more than boys, and increasing test scores by 0.25 standard deviations more. Girls also prove more sensitive than boys to changes in the distance to the nearest school. So much so that providing a community based school virtually eliminates the gender gap in enrollment (from 21 percentage points in control villages) and reduces the test score disparity by a third after a single year.

[1] This project could not have been completed without the assistance of many individuals. We appreciate the help and patience of the CRS team in Afghanistan while collaborating with us. Nicole Rigg and Nathan Falkner provided valuable research assistance. We are particularly indebted to Matthew Hoover for his exceptional work as both a research assistant and project manager. Finally, we also thank Dr. Saeed Mahmoodi for his work as our survey manager. This project was supported by grants from the National Science Foundation, the Spencer Foundation, and the Weikart Family Foundation. Valuable comments were provided by Janet Currie, Bentley MacLeod, and Miguel Urquiola as well as the seminar participants in the Columbia University School of Social Work Family Demography and Public Policy Seminar.

I. Introduction

Despite calls for universal primary education and an end to gender disparities in education, developing countries have not yet succeeded in either goal. Too few children go to school and of those who do not go, girls constitute the majority. In 2006, 73 million school-aged children were not enrolled in primary school. This is up from 103 million in 1999, but still below the millennium development goal of full primary school enrollment by 2015 (UN, 2008a). Unfortunately, over half of these children are girls (55 percent) and countries progress towards the goal of eliminating the gender disparity in primary education is mixed. Of the 113 countries that failed to meet the Millennium Development Goals preliminary target for gender equity in 2005, only 18 are considered by the UN to be "on track" to meeting the goal by 2015. The worst performers are countries in Sub-Saharan Africa, Oceania, and Western Asia, the setting for this study (UN, 2008b).

These challenges are particularly acute in rural areas. Globally, 25 percent of children in rural areas are out of school compared with 16 percent of children in urban areas (UN, 2008a). The same is true for gender equality. In urban areas, boys and girls, on average, attend school at similar rates. However, in rural areas gender disparities persist (UN, 2008c).

The major challenge is to determine whether or not these figures reflect a supply side or demand side challenges. On the supply side, educational institutions are often scarce in rural areas. Schools are inaccessible for many families and even when accessible children may have to travel long distances – sometimes walking as much as 10 kilometers – to attend (Al-Qusdi, 2003; Kristiansena and Pratiknob, 2006; Adele, 2008). Attending school in such conditions requires significant investments in time, transportation, or alternative housing strategies like arranging for children to stay with relatives for periods of time. On the demand side, rural areas usually offer few opportunities for skilled employment possibly reducing the returns to education. In addition, the opportunity cost of children's labor may be high, particularly in agricultural communities (Jafarey and Lahiri, 2005; Lloyd et al., 2005; Schultz, 2004; USDOL, 2000).

Both of these causes could also explain the existing gender disparities. On the supply side, the costs of girls traveling long distances may be more than boys because families may consider girls to face unique risks both to safety and chastity (Lloyd et al., 2005; Sutton, 1998). In addition, existing educational infrastructure may be better suited to meet the needs of boys rather than girls. For example, lack of separate sanitation facilities, female teachers, and gender-segregated classrooms are considered to be a greater deterrence to girls' enrollment than boys (UN, 2008c; Al-Qudsi, 2003; Kristiansena and Pratiknob, 2006; Adele, 2008).Families also have gender specific preferences over their children's behavior that generates potential demand side differences. Girls are much more likely than boys to perform domestic chores like cooking and child care, possibly making the opportunity cost of their time higher than their brothers' (Sutton, 1998). Additionally, the return to girls' education may not be as high as that of boys. Girls may marry early and be engaged solely in managing the family, limiting the returns to education to the rearing of children. Even if they do eventually enter the labor force, the industries in which most women work pay lower wages than men typically earn (UN, 2008c) and income generated usually benefits the family of the husband, creating an externality from the perspective of the girls' family.

In this study, we focus on a single major supply side challenge – proximity of schools. Unfortunately, governments do not randomly choose the location of schools, and any deliberate placement of school is likely to be correlated with the outcomes of interest: school participation and test scores. The results are that simple correlations between the distance a child has to travel to school and their participation rates are likely to be biased. For example, governments could place schools in areas with high demand for education in an effort to meet that demand, generating a positive bias. Alternatively, in areas with low school participation, governments might choose to increase the supply of educational resources in the hopes of increasing the number of enrolled students, potentially creating a negative bias. Or governments might simply place schools in areas with the densest populations, which would have an ambiguous effect.

We create exogenous variation in the distance that children have to travel to attend formal classes by conducting a randomized evaluation of a program designed to minimize the distance that children must travel to go to school by starting schools directly in children's villages. With a sample of 31 villages and over 2,000 children between the ages of 6 and 11 in two districts in northwest Afghanistan, we randomly assigned 13 villages to receive community-based schools a year before the schools were supplied to the entire sample. This phased-in approach allows us to estimate the one-year impacts of the community-based schools on girls' (and also boys') school attendance and knowledge of math and the local language, Dari. The experiment also allows us to randomly vary the distance that children must travel to attend schools providing an opportunity to measure the causal effect of distance on school participation and test scores.

Geographic proximity has a dramatic effect on children's academic participation and performance and has tremendous potential for reducing existing gender disparities in rural areas. The intervention, however, is extremely effective. The presence of a community-based school increases overall enrollment in formal schools (a set that includes community-based schools, government schools, and madrassas) by 42 percentage points, and increases test scores by 0.5 standard deviations overall. Those students attending a formal school experience an increase in test scores of 1.2 standard deviations. Distance proves particularly important for school enrollment with enrollment rates and test scores falling by 15 percentage points and 0.19 standard deviations for every mile a child has to walk to school.

These benefits, however, accrue disproportionately to girls. Their enrollment rates increase by 23 percentage points more than boys and their overall average test scores increase by 0.25 standard deviations more. The test score effects on girls who attend a formal school is the same as that for boys, implying that the larger impact on overall test scores is solely due to higher enrollment rates. Girls are also more sensitive than boys to distance. The enrollment rate for girls falls by 19 percentage points per mile while boys' enrollment only falls by 13 percentage points; girls' test scores fall by 0.24 standard deviations per mile while boys fall by 15 percentage points. The net effect of these disproportionate impacts is that in the treatment group the enrollment gap between boys and girls is almost eliminated, falling from 21 percent to 5 percent, and the test score gap is reduced by a third after only a single year.

This study builds upon a growing literature that investigates both the effects of reducing the upfront costs of education and that explore the family decisions processes surrounding the decision to send children to school. Most directly, there is a rich literature focusing on the role of financial incentives in decision to participate in school This includes the pioneering evaluation of the PROGRESA program that documented the efficacy of conditional cash transfers as a mechanism for encouraging enrollment and attendance (Schultz, 2004). These results have been replicated by many other researchers in many other countries (Cardoso and Souza, 2004; Levy and Ohls, 2006; Barrera-Osorio et al., 2008; Filmer and Schady, 2008 among many others). A distinct but related literature has documented that reducing user-fees can similarly increase enrollment (Barrera-Osorio, Linden, and Urquiola, 2007; Borkum, 2009).

A smaller, but growing literature has begun to investigate the intra-household and peer decision processes that determine enrollment patterns. This includes Berry (2009) who investigates the role of child-parent bargaining around school enrollment, and Bobonis and Finan (2007) and Lalive and Cattaneo (2006) who use the original PROGRESA experiment to show peer externalities from the higher levels of enrollment by treated students. Angelucci, De Giorgi, and Rasul (2007) follow a similar strategy, demonstrating positive externalities to the extended family members of beneficiaries. Studies of the intra-household externalities on the siblings of beneficiaries show mixed results. Filmer and Schady (2008) find no overall effect on the siblings of beneficiaries while Barrera-Osorio et al (2008) find a similar affect for most siblings and a negative externality on academically motivated siblings.

The paper is organized as follows. Section 2 provides a description of the state of education the Afghanistan and a description of the intervention used as the treatment in this study – the PACE-A community-based school program. Section 3 outlines the research design including the models used in the analysis. Section 4 describes the data we collected and provides a description of the sample of children under analysis. In section 5, we assess the internal validity of the study, and in Section 6 we analyze the effects of the program estimating first the reduced form effects of the intervention, then the effects of distance on children's school participation, and finally, we disaggregate the results by gender. Finally, we conclude in Section 7.

II. Afghanistan and the PACE-A Program

A. Education in Afghanistan

Afghanistan's educational infrastructure has been crippled by decades of war. Most notably, the country was ruled by the Taliban from 1996 until the latter half of 2001, a group that is openly opposed to the education of girls and women. As of 2007, half of school aged children were un-enrolled, and of primary school aged children, only 37 percent attended school (UNDP, 2007). As expected, there are significant differences between the experiences of boys and girls. Of those students who were educated, boys had an average of eleven years of education while girls had an average of four (UN, 2009). Of currently enrolled students, girls make up only a third of the student population (MoE School Survey, 2007). Despite these challenges, the government has had continuing success in expanding the reach of educational services. In 2001, only 900,000 children were enrolled in school, but by 2007, this number had reached 5.4 million (UNDP, 2007). In 2007 alone, 800,000 new children were enrolled and girls made up forty percent of these students (MoE School Survey, 2007).

One of the major challenges is the provision of educational services in rural areas. Distances between villages can be great and traveling between them can be dangerous – especially for young children and girls. As a result, despite the fact that eighty-three percent of schools are in rural areas, urban schools serve a disproportionate share of the student population – thirty-five percent of students as opposed to 65 percent of children in rural schools.

Afghanistan currently has four types of educational institutions. Three types of educational institutions provide a formal education[2] designed to be the primary source of the education for children. The first type of formal institution is public schools run by the government (here after referred to as "government schools"). These are usually multi-room schools designed to serve large numbers of children and in rural areas, to serve children from multiple villages. These schools face most of the same challenges as government schools in the rest of the developing world, including high rates of teacher absenteeism, a shortage of teachers, insufficient resources, and poorly maintained infrastructure (Adele, 2008; Alcazar et al. 2006).

The second type of formal institution is a madrassa. This is a formal religious school that provides a religious-based education (though religious content is also taught in government schools), often including the official government curriculum as well. The final type of formal institution is community based schools. These are village-based schools designed to serve only an individual village. They have been one of the major educational interventions supported by international aid agencies with the goal of increasing exposure to the official government curriculum in rural areas – particularly among girls who are reportedly less able to travel to schools outside of their villages than their male peers. While these schools are typically managed by international development organizations, the long-term goal is to integrate them into the national education system (MoE Progress Report, 2007). Currently, the government school system serves the vast majority of enrolled students, ninety-five percent, while madrassas and community-based schools serve two and three percent respectively. Finally, most Afghani villages also have mosque schools associated with the village's mosque. These schools provide only supplemental religious education.

[2] We will refer to these institutions that teach the formal Afghani government curriculum as "formal schools". The point is to distinguish them from the mosque schools that exist in each village and which provide only supplemental religious education and are more analogous to Christian Sunday schools.

Our study takes place in Ghor province in the Northwest section of the country. Compared to the south and southeast parts of the country, this area is much more secure and more stable. While there is some military activity related to the Taliban insurgency, tribal conflicts and inter-village rivalries seem to be a more significant source of insecurity. That said, the enrollment rates look similar to those in other rural parts of the country. Of children aged six to thirteen, only twenty-eight percent are enrolled in school. The gender gap in enrollment is almost seventeen percent – with thirty five percent of boys and only eighteen percent of girls participating in school. One of the main challenges is distance – only twenty-nine percent of the population lives within five kilometers of a primary school (MRRD, 2007).

B. PACE-A: Providing Community-Based Education to Rural Areas

The Partnership for Advanced Community-based Education in Afghanistan (PACE-A) is a five-year, USAID-funded program to expand educational opportunities to children, especially girls, in areas of Afghanistan that lack formal governmental schools or where children lack access to governmental schools. The goal of the partnership is to expand learning and life opportunities to marginalized Afghan communities. To accomplish this goal, the PACE-A partnership relies on establishing community-based schools through its partners. This model of educational delivery has evolved from underground schools developed under Taliban rule and an earlier partnership known as the Afghanistan Basic Education Consortium, of which Catholic Relief Services (CRS) was a partner.

The partnership comprises four organizations: CARE (the primary grantee), the International Rescue Committee, Aga Khan Foundation, and CRS (our partner). The total value of the five-year partnership (2006-2011) is USD 24 million. Under PACE-A, each partner works with communities in their provinces of operation[3] to establish primary classes open to children in the community. The community agrees to provide the space for the community-based school, and originally, to provide compensation for the teacher of the school. As PACE-A has evolved, the Project Management Unit, the overall managing body of PACE-A, has worked with the Afghan Ministry of Education to include community-based school teachers on its teacher payroll throughout the country, subject to certain credentialing requirements. The community-based schools use the Ministry of Education curriculum in schools, and the PACE-A partner provides teacher and community training and support and materials for both teachers and students of the school. In some areas, due to adverse weather, additional supplies to winterize schools may be provided by the PACE-A partner.

Under PACE-A, each partner has specific responsibilities to provide educational materials for the community-based schools as well as provide trainings for teachers. Education materials include writing utensils, notebooks, books, and teacher materials. In addition, each organization also provides ongoing training for the teachers. Teachers received Project Management Unit-created training on topics such as monitoring and evaluation, teaching methods, and instruction. Initial trainings utilized Project Management Unit materials though subsequent trainings for teachers have used the Afghan Ministry of Education Teacher Education Program (TEP) materials. The purpose of using TEP as a training course is to streamline and certify community-based school teachers into the Ministry of Education system of educators.[4]

Within the community-based school classrooms, students are exposed to the government school curriculum in Afghanistan that students in public schools encounter. They study for a

[3] The provinces in which PACE-A currently operates are: Badakhshan, Baghlan, Balkh, Bamyan, Ghazni, Ghor, Herat, Kabul, Kapisa, Khost, Laghman, Logar, Maidan, Nangahar, Paktia, Paktika, and Parwan

[4] These benefits were also available to the members of the school management committee. However, in our sample, these committees were not functioning during the period of time in which the evaluation took place.

minimum of 2.5 hours a day, 6 days a week (excluding Fridays). Students receive instruction in Dari (language) and math. Teacher trainers and community mobilizers visit each community-based school at least monthly to assist teachers; field officers visit monthly to assess the functioning of the community-based schools.

PACE-A's presence extends into 16 provinces throughout Afghanistan, affecting over 2,500 individual communities. According to the Project Management Unit, the partnership will reach more than 90,000 learners, 60 percent of which are estimated to be girls. The primary object of the program is to improve access to education by reducing the distance that children must travel to attend school. By placing the school within the community, PACE-A hopes to improve children's participation, especially among girls. In addition, PACE-A hopes to improve the community structures that support these institutions and improve the quality of the education that they provide. In the long term, the goal is to improve the ability of civil society to sustain the community-based education initiative and continue developing cooperation between the community-based education structures and the Ministry of Education in Afghanistan.

Our partner organization, CRS, operates in two provinces in Afghanistan. In the west, it operates in the districts of Guzara, Pashtun Zarghon, and Adraskan of Herat province. In the central plains, it operates in the districts of Chagcharan, Shahrak, and Tulak of Ghor province. CRS has committed to opening 204 CBS through the PACE-A partnership in total. The organization employs a team of people based in its Herat office and in field positions throughout Herat and Ghor provinces. The staff provides regular teaching and community support through a female/male pair of trainers and community mobilizers, coordination through a district representative to arrange logistics and training needs, and monitoring through field officers. In addition, from its Herat office, CRS arranges regular teacher trainings to credential its community-based school teachers with the Ministry of Education and school management training for community leaders to build community capacity to effectively manage the community-based school.

III. Research Design

A. Experimental Framework

The strategy behind a Randomized Control Trial is to generate exogenous variation in the treatment of interest, allowing for a direct estimation of the causal effect of the treatment on the chosen outcomes. In our study, we randomly assign the receipt of a community-based school to a subset of villages. The random assignment of the schools ensures that the receipt of the treatment is statistically independent of other child characteristics that might affect children's propensity to attend a school or their performance on our standardized tests. Effectively, the group of villages receiving a school should be similar in characteristics to those that do not and if this is true, then any average differences between the villages will be attributable to the allocation of the community-based schools.

Our initial sample consists of 34 villages in two districts (Sharahk and Chagcharan) chosen by Catholic Relief Services to receive schools as part of the PACE-A USAID-funded community-based education program. As part of the program, CRS had committed to placing a community-based school in each village over the two year period starting in the summer of 2007. We took advantage of the roll-out of the program to implement a phased-in experimental design in which a subset of the villages received the program in the summer of 2007 and the rest receive a school in 2008. The untreated villages in 2008 then served as a control group for the treatment villages that receive a school in 2007.

Due to logistical considerations and, at times, the political relationships between villages, CRS required that we not randomize the villages individually. Instead the villages were grouped into one of twelve groups of two to three closely located villages. Among these groups, five were selected as the treatment group that would receive the program in 2007 and seven were chosen to receive the treatment in 2008. The randomization was stratified by district. Unfortunately, shortly after the randomization, a conflict broke out that included one control group of villages in Sharahk comprising three villages. Due to the conflict, these villages were unapproachable and could not be surveyed. As a result, our final sample includes 31 villages, 13 treatment villages and 18 control villages.

Government schools are open through the end of the fall but close during the winter while the community-based schools in our sample remained open during the winter. As a result, we use two surveys to assess the intervention. One conducted in the fall of 2007 allows us to survey parents about their children's current educational status towards the end of the formal academic year. We also survey children in the spring of 2008 to collect a follow-up test after the end of the winter period when all of the community-based schools have finished their year. The details of the surveys are described more completely in the next section.

B. Econometric Models

We use four basic models in the analysis of the data. We use three basic models to compare directly the treatment and control villages. In addition, we use an instrumental variable's model to assess the effect of distance and school enrollment on test scores.

To compare directly the treatment and control groups, we first use a simple difference model:

$$Y_{ijk} = \beta_0 + \beta_1 T_k + \varepsilon_{ij} \qquad (1)$$

The variable Y_{ijk} in this specification is the variable whose average value is to be compared between the two groups for child i, in household j, and village k. The variable T_k is a dummy variable for whether or not village k was selected for treatment in the randomization process. The coefficient β_1 then provides the estimated differences in the variable Y_{ijk} between the treatment and control group. We use this specification primarily to compare the socio-demographic characteristics of the children in the treatment and control groups in Tables 1, 2 and 3, and to estimate the different between the two groups in enrollment and test scores in Tables 5, 6, and 7.

We also use Equation (1) to estimate the treatment effects, but we can improve the precision with which we estimate the treatment effects by controlling for socio-demographic characteristics that are unlikely to be affected by the treatment. This is done with the following specification which is also estimated through ordinary least squares:

$$Y_{ijk} = \beta_0 + \beta_1 T_k + \beta_2 X_{ij} + \varepsilon_{ij} \qquad (2)$$

The specification is identical to equation (1) with the addition of a vector, X_{ij}, of child and household characteristics. This specification is used in Tables 5, 6, and 7.[5]

[5] Many of the outcome variables that we consider are binary. In the tables, we exclusively use linear probability models. We have also estimated the main treatment effects using probit models which yield consistent estimates.

In Table 3, we compare the relative characteristics for children that attrit from the sample between the Fall 2007 and Spring 2008 surveys. To make this comparison we use a difference in differences estimator that compares the difference between attritors and non-attritors in the treatment group to the same difference in the control group. The model is estimated through ordinary least squares using the following specification:

$$Y_{ijk} = \beta_0 + \beta_1 T_k + \beta_2 Attrit_{ij} + \beta_3 Attrit_{ij} * T_k + \varepsilon_{ij} \qquad (3)$$

The variables Y_{ijk}, T_k, and ε_{ij} are all defined as in equation (1) and $Attrit_{ij}$ is a dummy variable set to 1 if student i in household j did not appears in the Spring 2008 survey. The coefficient β_3 in this model is then an estimate of the difference in characteristics between attritors and non-attritors across the treatment and control groups.

As we show below, a village's receipt of the treatment is correlated with outcomes that have an intermediate effect on student outcomes. For example, the primary purpose of the program is to reduce the distance that children have to travel to attend school. By placing a school within treatment villages, we exogenously reduce the distance between children and the closest available school. To measure the effect of distance on enrollment and test scores, we can estimate the following equations through instrumental variables:

$$\text{Enrolled}_{ijk} = \beta_0 + \beta_1 \text{Distance}_j + \beta_2 X_{ij} + \varepsilon_{ij} \qquad (4)$$

$$\text{Distance}_j = \delta_0 + \delta_1 \text{T}_k + \delta_2 X_{ij} + \upsilon_{ij} \qquad (5)$$

The variables Enrolled_{ijk} and Distance_j are respectively, a dummy variable for whether or not a child is enrolled and the distance to the nearest school. The remaining variables are then defined as in the previous equations. Equation (5) then provides the first stage regression for Equation (4), and the coefficient β_1 provides an estimate of the causal effect of distance on enrollment. This model is used in Table 8 and 9, and a variant of it in which the effect of school enrollment on test scores is estimated is used in Table 7. This latter estimate can be interpreted as a measure of the treatment effect on those students actually treated by enrolling in school, an estimate of the effect of the treatment on the treated.

In all of these models, it is important to take into account the correlation between children's performance and behaviors. Not taking this into account would cause us to underestimate the variance of the estimated treatment effect, resulting in over rejection of the null hypotheses at any significance level (Bertrand, Duflo, and Mullainathan, 2004). However, incorrectly specifying the level of correlation could cause us to overestimate the standard errors. The structure of the data in our sample is relatively simple with three possible groupings of children: the village group used for randomization, the village, and the household. To investigate the correlation in standard errors at each level, we estimated the treatment effect on our main outcomes of enrollment and tests scores using Equation 2 while including random effects at each level. Within this framework, allowing random effects at the village group and village level only explains between one and six percent of the residual variance while random effects at the household level explain about twenty percent. As a result, we cluster the standard errors at the household level.[6]

[6] In practice, the level at which the standard errors are clustered have little impact on the outcome estimates presented in the tables. Clustering at the village group level (with a correction for having 11 clusters (Cameron, Gelbach, and Miller, 2008)), for example, does generally increase the estimated variance of the estimates, but by only a small amount. The only difference is that the interaction between distance and gender presented in columns 5 and 6 of Table 9 are only statistically significant at the ten percent level, rather than the one percent level, and the difference in government school enrollment presented in column 4 of Table 6 and the TOT estimated effect of formal school enrollment on the Fall 2007 Dari test are significant at the five percent level, rather than the one percent level. All of the other outcome estimates are significant at the same levels. Given that there are only 11 village groups, we also bootstrapped the regressions with the primary outcome variables using a wild-cluster

IV. Survey Data

A. Survey Design

To assess the effects of the intervention, we conducted two waves of surveys in the fall of 2007 and then the spring of 2008. We designed the survey to fulfill four goals. First, in the absence of a baseline survey,[7] we collect information on socio-demographic characteristics that would not change as a result of the treatment, providing variables that we could use to compare the treatment and control groups to assess whether the randomization did indeed create comparable research groups. Second, we matched the data over time to create a panel data set. This allows us to compare attrition patterns between the treatment and control groups to ensure that differential migration patterns or other factors affecting the availability of households to complete a survey did not differentially affect the treatment and control groups. To assess the attitudes of parents towards school attendance, we administered a module to parents that asked questions about their preferences for children's school attendance. Finally, as our primary outcome variables, we asked parents about their children's school enrollment and directly tested children on their math and language skills.

The socio-demographic information collected in the first module of the survey was chosen such that each variable was unlikely to be affected by the treatment. This provides characteristics to assess the comparability of the treatment and control groups and to use as controls in regressions comparing the outcomes of students in the research groups. This included

---

bootstrap as recommended by Cameron, Gelbach, and Miller (2008), clustering at the village group level. The results are consistent with the small sample adjusted estimates.

[7] The original plan called for a baseline survey. However, less than halfway through that survey, the compound housing the staff members of our sponsoring organization was attacked by armed men. All work, including the baseline survey, was suspended until our sponsors could assess the basis for the attack and its implications for continuing work in the area.

the length of time that the family had lived in the village, the families' ethnic identity, the occupation of the primary earner, the primary earner's level of education, the family size, land holdings, and other similar characteristics. We collected the same information in both surveys to ensure that the same control variables could be used when analyzing data from either survey.

To help us match the data up across survey periods, we also collected information that would allow us to identify the households. This included the name of the head of the household and the longitude and latitude of each building. Geographic information was also collected on every government school and madrassa that a family reported their child attending as well as the community-based schools in the treatment villages. When combined with the household information, this allows us to measure the straight-line distance from each household to each educational institution in the sample.

The fall 2007 survey also contained a series of questions asking the household survey respondent about their preferences regarding their children's education. The questions were asked generically (rather than about specific children), and included questions about the difference that a formal education would make in a child's life, the subjects that parents' want their children to learn, and the ages at which children should stop going to school. In each case, the questions were asked separately of boys and girls in order to compare the responses for each gender.

The surveys contained two outcome modules with specific questions for each child between the ages of six and eleven within the household. The surveyor asked the household respondent for a list of all children in the age range targeted by the program. Then for each child, the surveyor collected information on whether the child attended school, the type of school attended, and the frequency of attendance.[8] The child's age, gender, and relationship to the head of the household were also collected.

Finally, for each child that was available to question directly, the survey administered a short test covering math and language skills. The questions were taken directly from the first grade government text books to ensure that the test covered material from the official Afghani curriculum. The math section included questions on number identification, counting, greater than or less than, addition, and subtraction. The language section covered Dari, the language taught in school, and included questions on letter identification, reading words of varying difficulty, basic grammar (subject verb agreement), and simple reading comprehension. The administered tests differed in that the spring 2008 survey covered a larger number of questions than the fall 2007 survey, though the same topics were covered in each survey.

Since this school participation information is self-reported, we were careful to assess the accuracy of this information (Barrera-Osorio, Bertrand, Linden, and Perez, 2008). First, the information itself is not obviously fabricated – not all parents report sending their children to school and the levels of reported school enrollment seem reasonable for the context and the information is consistent across the survey rounds. In fact the results are identical to government estimates of the average enrollment rates of boys and girls within Ghor Province (MRRD, 2007). In addition, we conducted qualitative semi-structured interviews with parents after the final survey was conducted. The information provided in those interviews was consistent with the information provided in the surveys. Second, the enrollment levels follow the patterns that one

[8] We also attempted to collect retrospective information on children's school enrollment. Unfortunately, families seemed to have difficulty distinguishing between different time periods. For example, the community-based schools were started in the summer of 2007, but a significant number of treatment families report sending their children to these schools in the spring of 2007. This is most likely due to our difficulty identifying specific historical time frames. Afghanistan uses both the Islamic and Western calendars though most of the families seem to describe time in terms of the passing of the seasons and the delineation between spring and summer may not have been as clearly and consistently understood as we had hoped. As a result, we restrict our attention to questions that relate to school participation at the time in which the surveys were actually administered.

would expect from the data, including, for example, higher rates of enrollment among boys and older children. Finally, the test scores in our data cannot be fabricated, providing two additional checks. We compare the relationship between demographic characteristics of the family and the probability of enrollment with the relationship between those variables and test scores and find that both measures follow the same pattern. Boys and older children attend school more often and also score higher on the exams. And second, for all of models, we estimate the effect on both enrollment and students' test scores – in all instances both measures provide consistent outcome estimates, showing both large effects on enrollment and learning levels.

B. Sample Size and Coverage

The survey was administered by a team of surveyors hired directly for the purposes of the study. The team comprised a single survey manager and 18 surveyors for the fall 2007 survey and 19 surveyors for the 2008 survey. The goal in each survey round was to survey all available households in our target villages. Each village has a readily distinguishable set of individual houses. These houses then served as the primary unit of analysis. Each house was identified and when the household was located, the team approached and surveyed the person most responsible for the family's welfare. As part of this initial survey, the surveyor created a list of eligible children between the ages of six and eleven. When possible, each of these children were then surveyed and tested.

Table 1 provides a tabulation of the responses from the survey comparing the coverage rates between the treatment and control groups. The information for the fall 2007 survey is provide in the first four columns and the information for the spring 2008 survey is provided in the last four columns. Households are listed in the spring 2008 survey regardless of whether or not they were surveyed in the previous survey round. In each case, we provide the total number of respondents in each category for each research group, followed by the difference between the treatment and control group and then the total number of respondents.

Along every category, the treatment and control groups are similar. Panel A provides the total number of households identified and the number actually surveyed. In both surveys about 93-95 percent of households were surveyed and coverage rates were similar across the two research groups. Only about two-third of surveyed households had children whose ages made them eligible to attend the community-based schools, and again the fraction of families meeting these criteria was the same across research groups in both rounds of the survey. In total, this provides a sample of 805 households in the fall 2007 survey and 794 in the spring 2008 survey. Finally, 1,490 and 1,477 eligible children from the fall 2007 and spring 2008 surveys respectively were identified in these households and had enrollment information provided to the surveyors. Of these, 1,374 from the fall 2007 survey and 1,401 from the spring 2008 survey were available to be tested.[9] As with the total number of households, the coverage rates are very high (92 and 95 percent) and balanced across the research groups.

C. Sample Description

Table 2 provides a snapshot of the villages in our sample absent the treatment. Using the data obtained from the control villages, the table contains regressions of the educational outcomes on the various demographic variables using the fall 2007 survey. This sample contains the 708 control children for which enrollment information is available and of which 653 took the test.

[9] Our initial sample included a small number of extremely large and wealthy households that we exclude as outliers. These included families with more than 20 household members, 10 units (jeribs) of land, or over 50 head of sheep. In each case, these families constituted the top one to two percent of households along each measure, and in total represented 3.3 percent of households in the fall 2007 survey and 2.9 percent of families in the spring 2008 survey. None of the results are sensitive to the exclusion of these households.

The children are almost equally divided between boys and girls (45.5 percent girls), and have an average age of 8.3 years. About 27 percent of these children report attending a formal school, very close the government's estimated province average of 28 percent (MRRD, 2007).

The first two columns contain regressions of formal school enrollment and overall test scores on child and household demographic characteristics. As expected the gender gaps in academic outcomes are very large. All else equal including distance to the nearest school, girls are 21 percent less likely to be enrolled in school than boys are. Possibly as a result, girls also score a shocking 0.69 standard deviations less than boys do on our standardized test. The enrollment gap is again close to the province average of 17 percent (MRRD, 2007). (We were unable to find similar province level statistics for children's test scores.)

Reassuringly, the enrollment measure seems to generally show the same pattern of correlations as the test scores which one would expect since formal school enrollment is the primary means of educating children in the official curriculum. As just mentioned, girls are less likely to be enrolled and score worse on the exam. Older children are both more likely to be enrolled and score higher on the exam. The same pattern is true for families with larger land holdings and larger numbers of sheep. And the opposite is true for families living farther away from the nearest formal school.

The last three columns contain a regression of the children's test scores on the fall 2007 exam on an indicator variable for whether a child reports attending a formal school and demographic characteristics. For each measure, formal school enrollment is correlated with higher test scores. The overall correlation is 0.56 standard deviations with a higher correlation for math (0.68 standard deviations) than for Dari (0.27 standard deviations). The lower correlation between enrollment and the Dari score is consistent with educational interventions usually having lower effects on language skills than math skills (see for example, Banerjee, Cole, Duflo, and Linden, 2004), and may also reflect the fact that children are taught Arabic, which shares many letters and words with Dari, by their local mullahs in the mosque schools. Interestingly, even controlling for school enrollment girls have lower test scores than boys.

Consistent with the conclusions of this study, the existing gender disparity among primary aged children seems inconsistent with the stated preferences of households. Figure 1 provides a histogram of the age at which households believe boys and girls should leave school. While it is clear that there is a preference to educate boys for a much longer period of time than girls, only a very small number of families believe that a girl should stop being educated prior to age eleven, the youngest students in our sample. The most common ages chosen for dropout of girls are between twelve and fifteen which corresponds to the typical age of marriage. This pattern is also consistent with the responses we received in the qualitative portion of our work that asked parents what role they saw for education in a child's life. For girls, families tended to emphasize the importance of education for the quality of a child's eventual spouse while for boys, families tended to emphasize work and financial support of eventual dependents.[10] These differences have obvious implications for girls' participation in school during the middle and secondary years, but they are inconsistent with the observed gender disparities for children of primary school age.

[10] In unpublished results (available upon request), we have compared the responses in the treatment and control groups to assess whether the proximity to a formal school may have changed parents' stated preferences. We find no differences.

V. Internal Validity

The purpose of a randomized evaluation is to ensure that the assignment of the treatment is orthogonal to other characteristics of the sample that may be correlated with school participation and test scores. Such correlations could arise in violation of the internal validity of our study in two ways. First, it is possible that the randomization simply created treatment and control groups with large differences in the characteristics of children that are also correlated with the outcomes of interest. Second, even if the research groups are initially similar, it is possible that over-time the sample may be affected by processes that differentially change the composition of the two groups. For example, if the treatment group proved more mobile than the control group, we may lose more families due to migration. The net effect would be that while the groups were initially similar, differences would emerge over time that would comprise the study's internal validity. We measure the differences both in composition and changes in composition between the two surveys and find that the randomization succeeded in creating comparable groups of children.

To check for the similarities in the two research groups, we directly compare the average children in the treatment and control groups using socio-demographic characteristics that would not have been affected by the presence of a closer school. This is done in Table 3. Using the data from the fall 2007 survey, the first three columns contain estimates for all of the children in the sample while the second set of columns contains only those children who took the exam. The first column provides the average characteristic of the treatment group. The second column contains the control average, and the third column contains the difference estimated using equation (1). Panel A contains child demographic variables and Panel B contains the household characteristics.

On average all of the differences are practically small. A few of the differences are estimated precisely enough to be statistically significant, but even these are small in magnitude. Consider for example, the number of sheep owned by a household. On average, treatment families own about 7.6 sheep per household when considering the sample with all children while the control families only own 5.6 sheep, yielding a difference of 1.9 sheep per household. Two sheep is a relatively small difference, especially when we consider the relationship between the number of sheep and our outcomes presented in columns one and two of Table 2. The correlation is positive which may reflect the small difference in family wealth, but the estimated coefficient is only 0.008 percentage points per sheep. The difference of 2 sheep reflects a possible difference in enrollment rates between the two groups (absent the treatment) of 1.6 percentage points. Given that we estimate treatment effects in enrollment of 40 to 60 percentage points, it is unlikely that these small differences in the composition of the groups could significantly affect the estimated treatment effects.

Given the nature of the study, a particularly important variable to consider is the distance each child would have to travel to attend the nearest formal school. Using the geographic information we collected, we estimate the distance between each house and every formal school, excluding the community-based schools. In other words, we estimated the smallest distance children would have to travel absent the treatment for every child in the sample. The average differences are presented in the last row of Table 3. On average, children live about 3 miles from the nearest formal school, and the average difference in the distances between the children in each group is only a quarter of a mile. Figure 2 shows these differences graphically by plotting a non-parametric estimate of the density of distances for the treatment and control children. The distributions are very similar, especially when compared to the distribution after the treatment presented in Figure 3.

Table 4 investigates whether or not the sample of children we observe changes significantly over time. The first three columns compare the relative characteristics of children that attrit between the two survey rounds to those that do not. Panel A contains the raw attrition rates for each group. On average, the attrition rate is only about 16 to 17 percent with a difference of only 1 percent between the two groups. To check that the types of children in each group did not change significantly, Panels A and B contain comparisons of the children using child and household demographic characteristics considered in Table 3. Column one presented the simple difference in characteristics between attritors and non-attritors in the treatment group. Column two presents the same comparison in the control group, and column three presents the difference in these differences using equation (3).

There is no clear pattern in the relative characteristics of the attriting and non-attriting children in each of the research groups. In fact, the differences are of similar magnitude as those we observe between the treatment and control children in Table 3. This is consistent with the observations of our survey team that the main reason for failing to observe a family was not migration or other causes that could have a clear relationship to wealth or some of the other characteristics correlated with academic performance, but were rather causes that would be common to almost all of the families, like traveling out of the village temporarily to visit family members or go to a local market. As a result, the resulting difference in the characteristics between attritors and non-attritors of the treatment and control groups are similarly small.

The last three columns of Table 4 present the ultimate result of the attrition patterns by comparing the average characteristics of the children that appear in both the fall 2007 and spring

2008 surveys. Consistent with the small, inconsistent differences in the attrition patterns between the two groups, the differences between the non-attriting children are remarkably similar to the differences observed between all of the children in Table 3.

Finally, in results not presented in this draft, we performed similar comparisons to those in Table 3 using the data from the spring 2008 survey. We compared the characteristics of children for whom we had enrollment information to those students for whom we were able to obtain both enrollment information and test scores. And we compared the attrition patterns using just the sample of children who provided a test in both surveys. In all instances, the observed differences were as small as those presented in Tables 3 and 4.

VI. Outcomes

Given that the randomization created comparable treatment and control groups, the only major differences between the two groups is that the treatment group received a community-based school while the control group did not. As a result, we can attribute any difference in the groups in enrollment and test scores to the receipt of the treatment. We assess these differences in three steps. First, we assess the overall average, reduced form effects of the program on the children in treatment villages. Second, we use the exogenous variation in distance to estimate the effects of geographic proximity on children's enrollment and test scores. Finally, we take into account the gender of the children and compare the reaction of boys and girls to the intervention. Overall, we find that proximity is a significant determinant of children's academic achievement and that it plays an important role in ameliorating the existing gender disparities in our sample.

A. Overall Effects

We are primarily interested in the effects of the program on two outcomes: enrollment and test scores. Since the main purpose of establishing the new schools is to expose more children to the official Afghani curriculum, we first analyze the effects of the program on children's enrollment in schools teaching the formal government curriculum. Next, we disaggregate those changes in enrollment to assess the enrollment rates of children in the individual types of schools. Finally, we assess the differences in students' test scores using the tests administered both survey rounds. Along all dimensions, the program proves extremely effective.

Table 5 contains the main outcome of interest – enrollment in formal schools. Panel A includes all children while Panel B just includes those children giving a test in the respective survey round. For each panel we first estimate the enrollment levels in: community-based treatment schools, formal schools during the fall of 2007, and formal schools in the winter of 2008. The first column contains the average enrollment rates for the treatment group followed by the average enrollment rates for the control group, and then the simple difference in these rates estimated using equation (1). The final column contains the average differences controlling for socio-demographic characteristics using equation (2).

The program has a large impact on student enrollment. Turning to the first row of Panel A, the program causes 56 percent of children to attend the community-based schools, reflecting the fact that even when a school is readily accessible not all children attend. Some of these children switch from having attended other schools, and the resulting overall increase in enrollment rates for treatment schools is about 47 percentage points – a very large increase in enrollment over the control enrollment rate of 27 percentage points. The final row shows that formal school enrollment continues into the winter for the treatment group, emphasizing that an advantage of the community-based schools is a more flexible academic year. Government schools close for the winter because moving between villages is almost impossible given the lack of roads and other infrastructure for managing the very heavy snowfalls. Panel B shows that there is no difference in the results for students who also took the test which is consistent with the fact that there are no systematic differences between tested and non-tested students.

Comparing the estimates in column three with those in column four, the estimates from the simple difference estimator (equation (1)) and those from the estimator that controls for demographic characteristics (equation (2)) are very similar. This similarity reinforces the conclusions of Section V that the treatment and control groups are very similar in the observable characteristics in our data. Had they been significantly different along dimensions correlated with student enrollment, the point estimates for the two estimators would differ significantly.

The ultimate effect of the community-based schools is a high level of primary school enrollment given the control average of 27 percentage points. While all children still do not attend school, an impressive 74 percent of children do. This is still below the Western Asia average of 2006, but it is equal to the global rural average enrollment rate in 2006 (UN, 2008c). This suggests that community-based schools could be an important tool for achieving universal primary education in rural areas.

In Table 6, we disaggregate these estimates and estimate changes in the enrollment of students in the individual types of schools. The layout is similar to that of Table 5, except that we only present the results for all children in the data set.[11] The first row provides the change in community-based school enrollment for reference (this is the same estimate as row one from Panel A of Table 5). The second row, estimates the difference in government school enrollment, and as one would deduce from the difference between community-based school enrollment and

[11] Estimates for the sample of children who also took the test are similar.

formal school enrollment, enrollment in government school falls by 10 to 15 percentage points. The last two rows estimate changes in children's participation in the informal mosque schools that teach a Koranic curriculum. Interestingly, despite the fact that the curricula are different, the community-based schools reduce enrollment in these informal schools by 8 to 17 percentage points depending on the season. It may be that some families view this type of informal education as a substitute for formal education, sending their children when it is the only option available, but withdrawing their children when more formal schooling options are available.

Finally, Table 7 presents the difference in average test scores between the treatment and control groups on both of the exams. Columns one through four are organized as in Table 5 and 6 with Panel A providing the results for the fall 2007 exam and Panel B providing the results for the spring 2008 exam. Starting with Panel A, the results demonstrate the school generated large changes in students' test scores. On average, the program generated an overall change in test scores of 0.59 standard deviations, a result that is statistically significant at the one percent level. The change in math scores was larger than the improvement in language scores (0.62 standard deviations versus 0.42 standard deviations respectively), but both are quite large. The relative pattern of results is consistent with those of other interventions that generally find larger effects of treatments on math scores than languages scores (see for example Banerjee, Cole, Duflo, and Linden, 2004). The results for the spring 2008 exam in Panel B show the same patterns.

These estimates in columns one through four are the overall average estimates for all children in the villages. However, as shown in Table 5, not all children in the treatment villages attend a formal school (either a community-based school or otherwise) and some of the control students attend formal schools. Both of these mean that the overall average treatment effect, or the intent-to-treat effect underestimate the actual change in test scores on treated children due to the treatment. To estimate the treatment effect on those children who are actually treated as a result of the program, we use an instrumental variable procedures similar to the one using equations (4) and (5) in which we use the relationship between formal school enrollment and the treatment assignment as a first stage for a regression of students' test scores on formal school enrollment. These treatment-on-the-treated estimates are presented in column five. As expected they are much larger than the average treatment effects. One year of formal school causes an increase in test scores of 1.2 standard deviations – an extremely large increase in scores.

B. The Effect of Distance

Because the main variation generated by the treatment is the distance that children must walk to school, we next use this variation to estimate the relationship between the distance to the nearest school, enrollment, and academic performance. Using the geographic coordinates of every school attended by a child in our data set as well as the coordinates of each household and each community-based school,[12] we calculate the distance between each household and the nearest formal school of any type. This is the same measure of distance used in the last row of Tables 3 and 4, except that this measure includes the location of the community-based schools. We then estimate the relationship between enrollment and test scores and distance by isolating the distances correlated with the treatment assignment using two-staged least squares with equations (4) and (5).

An important consideration is whether or not the community-based schools differ significantly in other ways from the other formal schools. If there are significant differences other than distance that are correlated with students' academic performance or parents' decisions

[12] Note that the locations of the community-based schools were fixed during the recruitment period because each village had to demonstrate the existence of a suitable location for the presence of a school in order to be eligible for the PACE-A program.

to send their kids to school, then this specification will attribute the differences in enrollment and test scores generated by those characteristics to geographic proximity. If parents, for example, were attracted to government schools because they were more established institutions with a more traditional structure, those preferences would cause us to underestimate the relationship between distance and education. As described in Section 2, however, the two sets of teachers teach the same curriculum and experience a common training program. In fact our exams suggest that students benefit equally from both types of schools and that if anything, community-based school students score slightly higher than those attending other formal schools.[13]

To gauge the effect of the treatment on distance, Figure 2 shows a non-parametric estimate of the density of household's distance to the nearest non-community based school. The distributions without the treatment are very similar. As noted in the last row of Table 3, the average difference in distance is only a quarter of a mile. Figure 3 shows the impact of the treatment. These estimates contain similar regressions, but measure the distance of each household to the nearest formal school including the community-based schools. The difference is dramatic. Almost no treatment households have to walk more than 2 miles to the nearest school – the average distance to the nearest school falls from 2.9 miles to 0.3 miles. Column one of Table 8 shows this comparison as the first stage regression (equation (4)) in the instrumental variables framework. Controlling for other observable characteristics, the treatment reduced the distance families need to travel to the nearest school by 2.7 miles.

The effect of having a closer school is dramatic. Figure 4 depicts a non-parametric regression of the probability of enrollment in a formal school as a function of the distance to the

[13] To estimate this, we estimate an OLS regression in which the total score on our exam (in the Fall of 2007) is regressed on demographic characteristics and dummy variables for attending community-based schools or other formal schools. The coefficient on attending a community-based school and other formal school are 0.95 and 0.81 standard deviations respectively, both statistically significant at the one percent level.

nearest school. Within a mile, enrollment rates are very high – above 70 percent. While below this level, the enrollment rates begin to decline quickly in distance until the enrollment rate is around 30 percent for children living more than 2 miles away from school. Figure 5 shows the same relationship, except using total test score rather than enrollment. The relationship generally follows the same pattern with test scores within 1 to 1.25 miles averaging about 0.6 standard deviations and then scores declining dramatically to about 2 miles out.[14]

Columns two through five of Table 8 estimate this relationship within the instrumental variables framework using equations (4) and (5). For reference, column (2) provides an estimate of the OLS estimate of equation (5). The coefficient is already negative and about thirteen percentage points – larger than the correlation observed using the control group alone (Table 2 column one). Column three contains the instrumental variables estimate of the relationship. The coefficient suggests that children are very sensitive to distance with enrollment declining 15.9 percentage points for each mile, consistent with the dramatic reduction in enrollment depicted in Figure 4. The estimate is significant at the one percent significance level. Column four contains a similar regression, however, only using those children who took the exam. The coefficient is almost identical (-0.16) and also statistically significant at the one percent level. Finally, column 5 estimates the same relationship using the score on the standardized test as the depended variable instead of enrollment, and the results are consistent with the enrollment regressions. The coefficient suggest that test scores fall by 0.19 standard deviations per mile that children have to travel; this means that having to travel a mile less to school has an equivalent effect of children's test scores as many of successful classroom based interventions (see for example,

[14] The rise in test scores after 2.5 miles is consistent with a small increase in enrollment in the same range. As we will show below in Section VI.C., this is entirely due to the behavior of boys, and may reflect families far enough away from the nearest school finding alternative strategies to send some of their children to school.

Banerjee, Cole, Duflo, and Linden, 2004; Muralidharan and Sundararaman, 2008; He, MacLeod, and Linden, 2008).

C. Effects by Gender

Given the strong overall impact of the program, the existing gender disparities, and presumed relationship between gender and the effects of proximity, we next assess the differential effects of the program on boys and girls. These results are presented in Table 9 and Figures 6 and 7. We duplicate all of the preceding outcome estimates allowing for interaction with the gender of the child. All of the estimates demonstrate that while boys are affected by proximity, girls are much more sensitive to distance than boys. The difference in sensitivity by gender results in the virtual elimination of the gender gap in enrollment in the treatment villages and a reduction in the test score gap by a third after only one year of treatment.

Columns one and two of Table 9 replicate the enrollment effects estimated in Table 5 but include an interaction of the treatment indicator with an indicator for a child being female. In both cases, the treatment has larger effects on girls than boys. Turning first to the basic treatment variable of enrollment in a community-based school, the average treatment effect for boys is 46.5 percentage points, but the treatment effect for girls is almost fifty percent higher – 69 percentage points, a difference that is statistically significant at the one percent level. As we noted before, children may have already been attending school. So, we consider our primary outcome variable, formal enrollment. Here too girls react more strongly than boys. Boys increase their enrollment in formal schools by 34.9 percentage points while again girls increase their enrollment by 44 percent more for an increase of 50.2 percentage points which is again significant at the one percent level.

The change in test scores is also larger for girls than for boys. Column (3) estimates the treatment effect using the fall 2007 test. The increase in test scores for girls is significantly larger than that for boys – 0.63 standard deviations versus 0.38 standard deviations. Girls score 0.25 standard deviations more. Column four estimates the effect of the treatment on the treated as the last column of Table 7. Here girls that attend a formal school due to the intervention do seem to score a bit higher than boys, but the difference is smaller than the other estimates and not statistically significant at conventional levels – suggesting that much of the larger effect on girls test scores is due to the larger number of girls participating in a formal school rather than girls benefiting more from formal schooling than boys.

Given the larger changes in enrollment, it seems reasonable that girls may be more sensitive to distance than the boys. To estimate this relationship, we recreate Figures 4 and 5, but divide the sample by gender. Figure 6 contains the relationship between enrollment and distance. As with the overall results, both boys and girls are less likely to attend the further a child has to travel to go to school. However, the slope of girls' enrollment in distance is much steeper than the boys. For schools close to a child's home, children of both genders are almost equally likely to attend. However, 1.5 miles out, there is already a gap of almost 10 percentage points with girls attending less than boys. Figure 7 contains the estimated relationship between test score and gender. Like the previous figure, all children's test scores decline with distance, but again, the slope of girls' scores in distance is more negative than that of boys. The difference in scores for girls is remarkable – girls close to school score about 0.4 standard deviations above the mean while those more than 2 miles away score about 0.3 standard deviations below the mean.

Columns five and six of Table 9 replicate the instrumental variables estimations from Table 8 allowing for an interaction between distance and gender. The results support the dramatic results presented in Figures 6 and 7. Column five estimates the effects on school enrollment. Again, boys are sensitive to distance – reducing enrollment 13.2 percentage points for each mile. However, girls are much more sensitive to distance, reducing their enrollment by an additional 5.8 percentage points per mile, a difference that is again statistically significant at the one percent level. Column six presents the results of the same specification but with test scores rather than enrollment. The results are just as dramatic. Boys test scores decline by 0.148 standard deviations per mile, but girls test scores decline by an additional 0.088 standard deviations.

Given the significant differences in children's reaction to distance, we re-estimate the gender disparities in the treatment villages. To do this, we estimate the same equations we estimated in columns one and two of Table 2, but estimate them on the sample of children in the treatment villages. The results dramatically illustrate the importance of proximity in explaining the existing gender gap. While the gender disparity in enrollment was 20.8 percentage points in control villages, the difference for treatment villages is only 4.0 percentage points, a difference that is not even statistically significant at conventional levels (p-value = 0.192). The test score disparity is also significantly lower in the treatment villages. While the disparity in the control villages is 0.69 standard deviations, the disparity in the treatment group is 0.45 standard deviations – a difference of over a third.

VII. Conclusion

The results of this paper prove that geographic access – proximity to school – is a critical component in both improving primary school enrollment and ameliorating the enrollment gap in primary education. We show that a program designed to place formal schools within villages has a significant effect on children's school participation and learning. The program increases enrollment in formal schools by 47 percentage points over villages not receiving a community-based school. Average test scores among all primary aged children increase by 0.59 standard deviations with improvement in both math and language skills, with larger increases in math than language. The scores of the children that actually enroll in school as a result of the program improve by 1.2 standard deviations. The purpose of this program is to reduce the distance that children have to travel to study the official government curriculum. The changes in distance significantly affect children's outcomes. On average, children's enrollment declines by 16 percentage points for every additional mile that a child has to travel to school. Children's test score suffer similarly, decreasing by 0.19 standard deviations for every mile.

Proximity is particularly important for girls. Placing a community-based school in a village increases girls' formal school enrollment by 15 percentage points more than boys. On average, the test scores of all girls in the village also increase by 0.25 standard deviations more than boys, though this is primarily due to higher enrollment rates rather than girls' benefiting more than boys from school participation. Girls are also much more sensitive to distance. The enrollment of boys falls by 13.2 percentage points for every mile while the enrollment rate of girls falls by an additional 5.8 percentage points. Similarly, boys test scores fall by 0.15 standard deviations per mile while girls' scores fall by an additional 0.09 standard deviations per mile. The net effect of these differences is that placing a school in each village dramatically reduces the existing gender disparities. The community-based schools eliminate the enrollment gap between boys and girls with a difference of only 4 percentage points compared to a 21 percentage point deficit in the control villages. The test score gap also falls significantly – the gap in treatment villages is over one third less than in control villages after only a year of treatment.

VII. Bibliography

Adele, M. E. (2008) “Afghanistan on the Educational Road to Access and Equity,” *Asia Pacific Journal of Education.* 28(3): 277-290.

Alcazar, L., Rogers, F.H., Chaudhury, N., Hammer, J., Kremer, M., and Muralidharan, K. (2006). Why are teachers absent? Probing service delivery in Peruvian primary schools. *International Journal of Educational Research*, 45, p. 117-136.

Al-Qudsi, S. S. (2003) “Family Background, School Enrollments and Wastage: Evidence from Arab Countries,” *Economics of Education Review*. 22: 567-580.

Angelucci, M., M. Rangel, and G. De Giorgi (2007) “Family Networks and School Enrollment: Evidence from a Randomized Social Experiment” Working Paper.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden (2007) "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 122(3): 1235-1264.

Barrera-Osorio, Felipe, Leigh L. Linden, and Miguel Urquiola (2008) “The Effects of User-Fee Reductions on Enrollment: Evidence from a Quasi-Experiment” Columbia University Working Paper.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. (2004) “How Much Should We Trust Difference in Differences Estimates?” *Quarterly Journal of Economics.* 119(1): 249-275.

Bobonis, Gustavo and Frederico Finan (2007). “Endogenous Peer Effects in School Participation” Manuscript. University of Toronto Department of Economics.

Borkum, Evan (2008) “Geographically targeted user fee elimination: Evidence on enrollment from South Africa’s no-fee schools,” Columbia University Department of Economics Mimeo.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008) “Bootstrap-Based Improvements for Inference with Clustered Errors,” *Review of Economics and Statistics*. 90(3): 414-427.

Cardoso, Eliana and André Portela Souza (2004). “The Impact of Cash Transfers on Child Labor and School Attendance in Brazil,” Vanderbilt University Working Paper No. 04-W07.

He, Fang, Leigh Linden, Margaret MacLeod (2008) "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program," Mimeo. Columbia University, Department of Economics.

International Child Labor Program (2000) “Chapter IV: Access to Primary Education,” *By the Sweat & Toil of Children: An Economic Consideration of Child Labor, Volume VI.* United States Department of Labor, Bureau of International Labor Affairs.

Islamic Republic of Afghanistan Ministry of Education (MoE) (2007). Schools Survey Summary Report. EMIS Department.

Islamic Republic of Afghanistan Ministry of Rural Rehabilitation and Development (MRRD) (2007). Ghor Provincial Profile. Retrieved 2/13/09 from http://www.mrrd.gov.af/NABDP/Provincial%20Profiles/Ghor%20PDP%20Provincial%20profile.pdf.

Jafarey, Saqib and Sajal Lahiri (2005) “Food for Education Versus School Quality: A Comparison of Policy Options to Reduce Child Labor,” *The Canadian Journal of Economics*. 38(2): 394-419.

Kristiensena, Stein and Pratiknob (2006) "Decentralizing Education in Indonesia," *International Journal of Educational Development*. 26: 513-531.

Lalive R, and A. Cattaneo (2006). "Social Interactions and Schooling Decisions," IZA Discussion Papers 2250, Institute for the Study of Labor (IZA).

Levy, Dan and Jim Ohls (2006). "Evaluation of Jamaica's Path Program: Final Report." Mathematica.

Lloyd, Cynthia, Cem Mete, and Zeba A. Sathar (2005) "The Effect of Gender Differences in Primary School Access, Type, and Quality on the Decision to Enroll in Rural Pakistan," *Economic Development and Cultural Change*. 53(3): 685-710.

Muralidharan, Karthik and Venkatesh Sundararaman (2008) "Teacher Performance Pay: Experiment Evidence from India," Mimeo, Department of Economics, University of California at San Diego

Schultz, T. P. (2004) "School subsidies for the poor: Evaluating the Mexican PROGRESA poverty program," Journal of Development Economics, 74(1), 199-250.

Sutton, Margaret (1998) "Girls Educational Access and Attainment" In *Women in the Third World: And Encyclopedia of Contemporary Issues*. Ed. Nelly P. Stromquist. 381-396. New York: Garland Publishing.

UNDP (2007). Afghanistan Human Development Report, 2007. Center for Policy and Human Development, Kabul University.

United Nations (2008a) Fact Sheet: *Goal2 Achieve Universal Primary Education.* United Nations Department of Public Information, Publication Number DPI/2517 H, September.

United Nations (2008b) *Fact Sheet: Goal 3 Promote Gender Equality and Empower Women,* United Nations Department of Public Information. Publication Number DPI/2517I, September.

United Nations (2008c). *The Millennium Development Goals Report 2008*. United Nations Department of Public Information.

United Nations (2009). Afghanistan Humanitarian Action Plan. United Nations.

Figure 1: Age at which Children Should Quit School

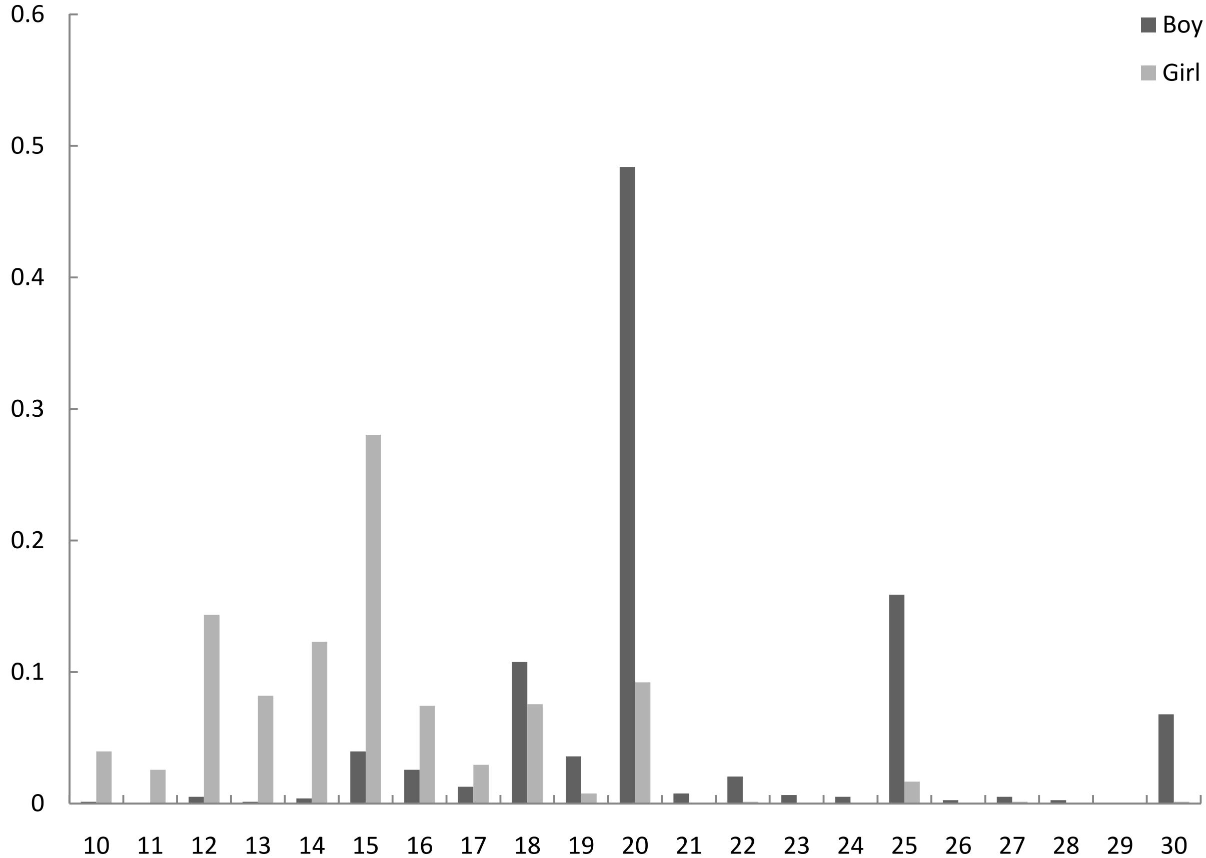

Figure 2: Density of Distance to Nearest Non-Community-Based Formal School

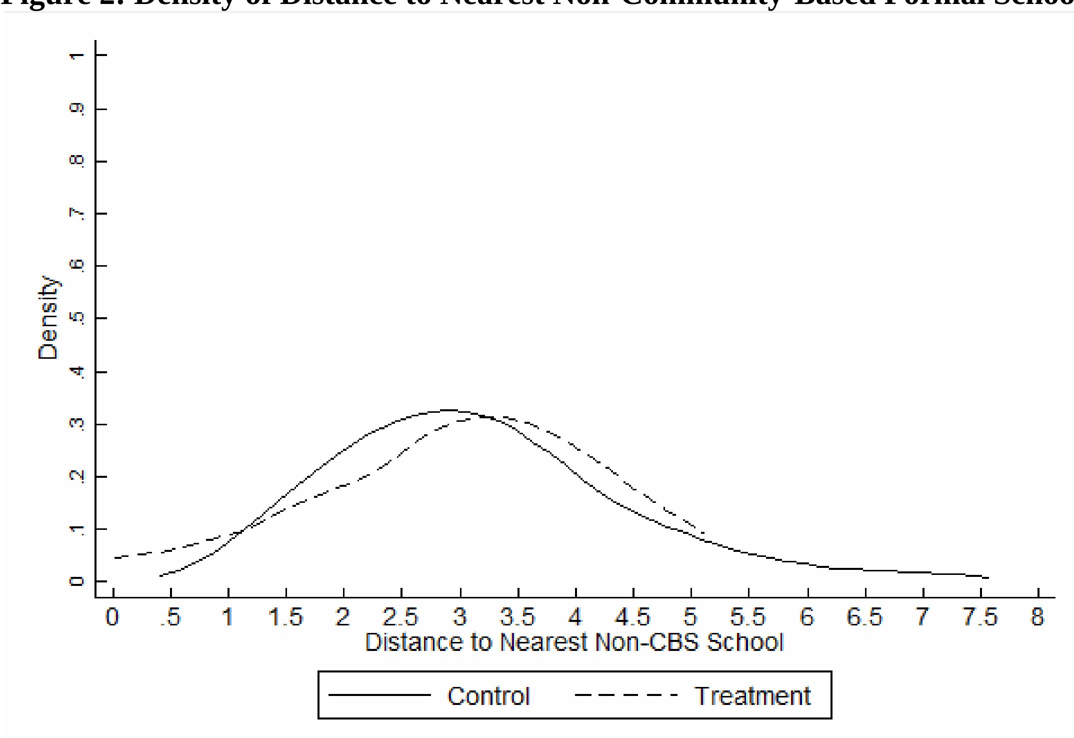

Figure 3: Density of Nearest Formal School

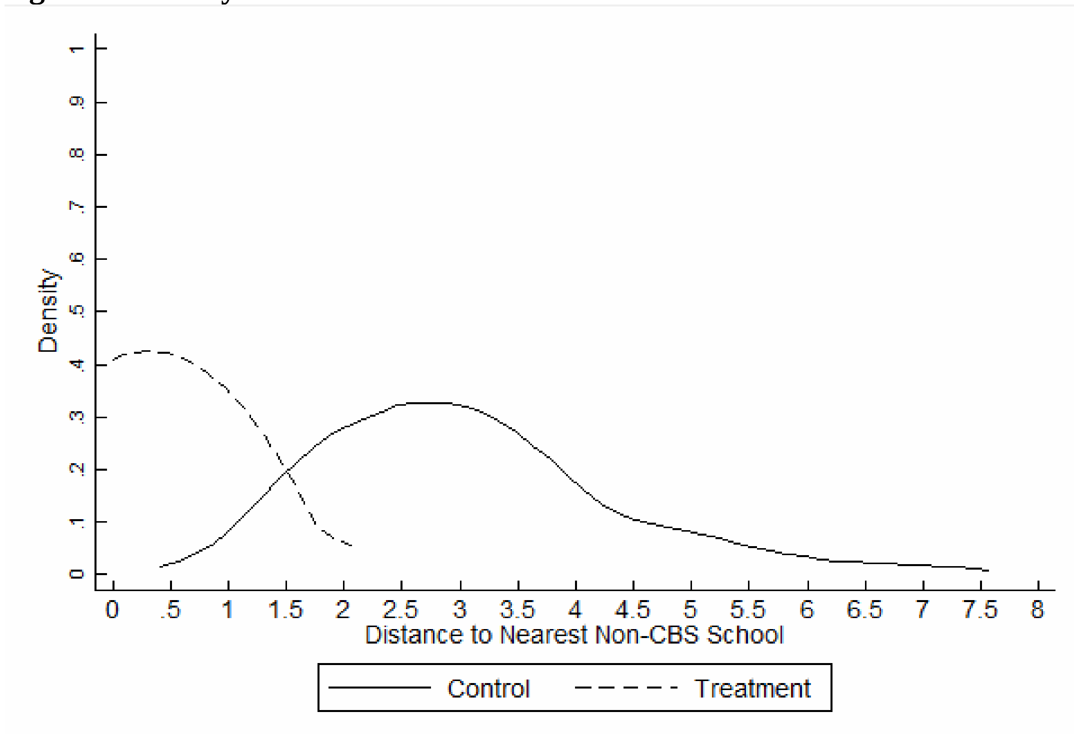

Figure 4: Enrollment as a Function of Distance to Nearest Formal School

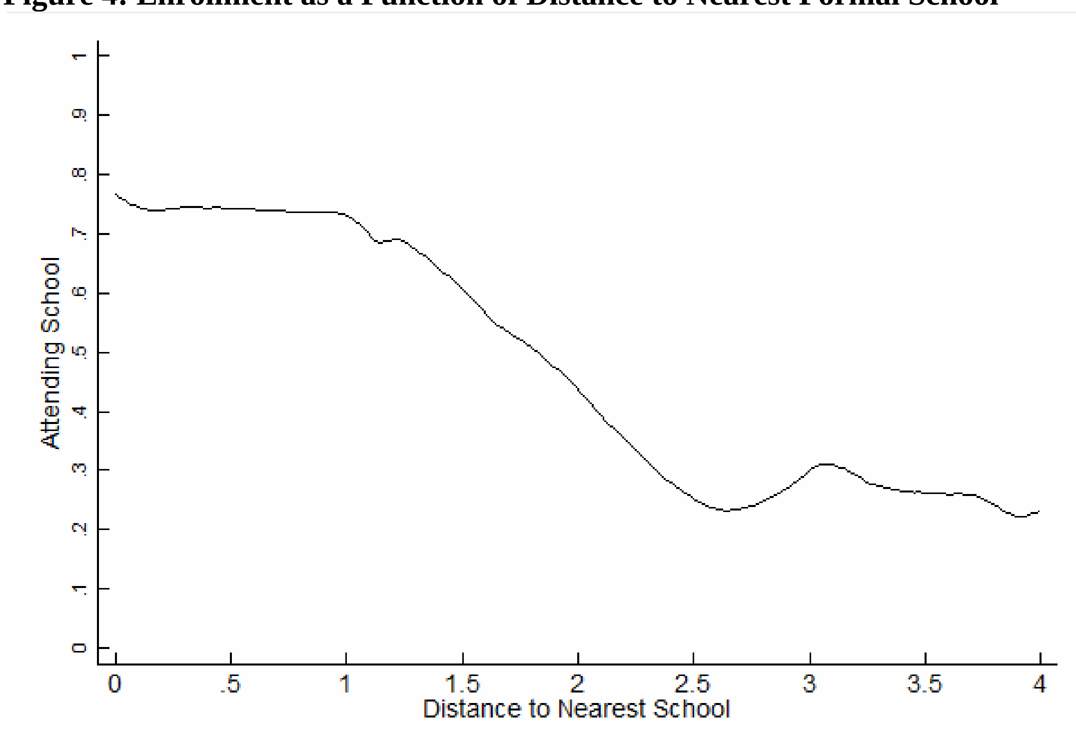

Figure 5: Test Score as a Function of Distance to Nearest Formal School

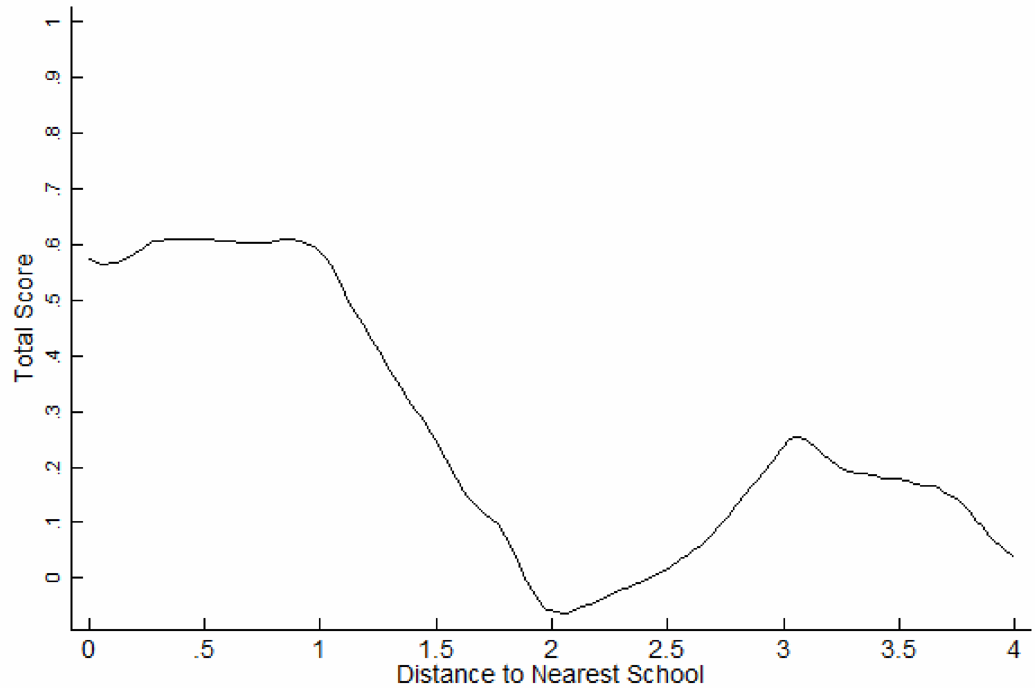

Figure 6: Enrollment as a Function of Distance to Nearest Formal School by Gender

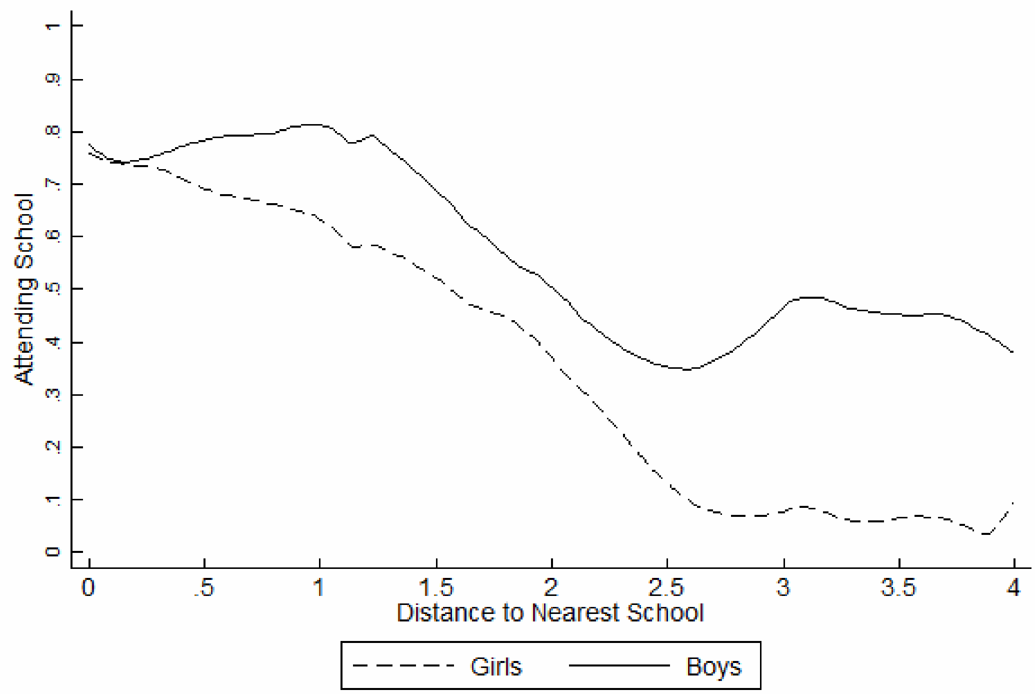

Figure 7: Test Score as a Function of Distance to Nearest Formal School by Gender

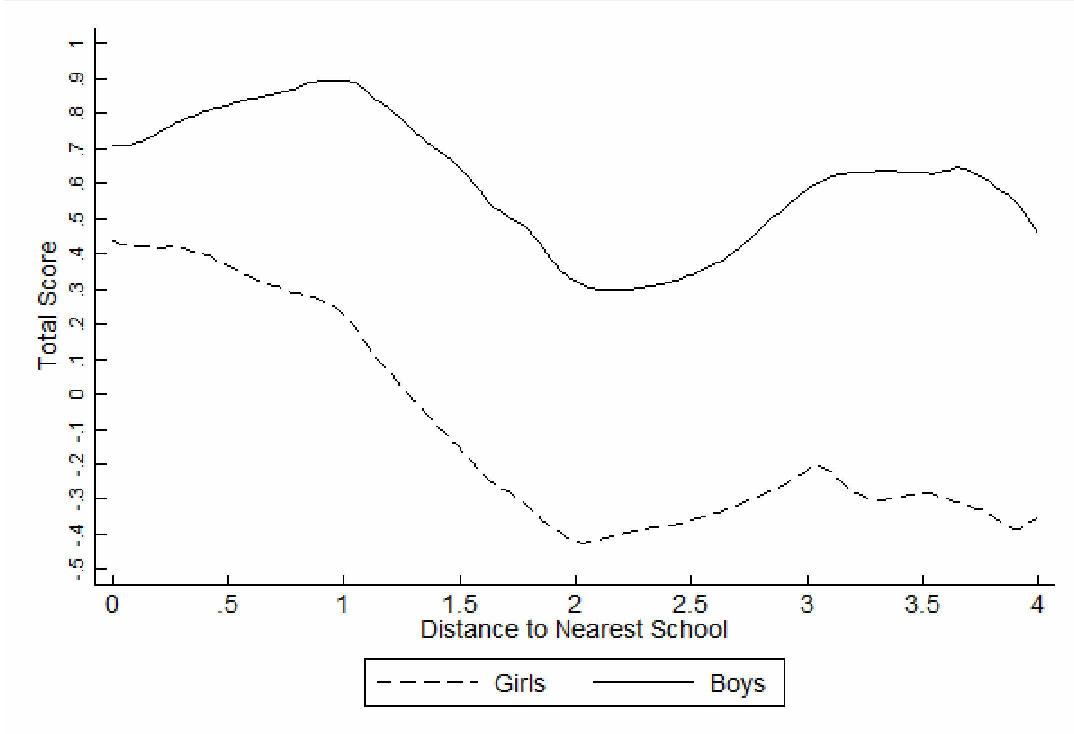

Table 1: Sample Size and Coverage Rates by Research Group

	Fall 2007 Survey				Spring 2008 Survey			
	Treatment Group	Control Group	Estimated Difference	Total	Treatment Group	Control Group	Estimated Difference	Total
Panel A: Households Surveyed								
Identified	680	663	17	1343	637	616	21	1253
Surveyed	635	628	7	1263	603	582	21	1185
Percent of Households Surveyed	0.934	0.947	-0.013	0.94	0.947	0.945	0.002	0.946
			(0.013)				(0.013)	
Panel B: Households with Children								
Households with Children	414	391	23	805	399	395	4	794
Percentage with Children	0.65	0.618	0.033	0.634	0.662	0.679	-0.017	0.67
			(0.027)				(0.027)	
Panel C: Children Tested								
Identified	782	708	74	1490	756	721	35	1477
Tested	721	653	68	1374	722	679	43	1401
Percent of Children Tested	0.922	0.922	< 0.001	0.922	0.955	0.942	0.015	0.949
			(0.016)				(0.011)	

Note: This table contains the tabulation of the sample used for the study divided by survey round and research group. The differences are estimated using equation (1) clustered at the household level. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 2: Correlation between Enrollment, Test Scores, and Demographic Characteristics in Control Group

	Enrolled (1)	Total Score (2)	Total Score (3)	Math Score (4)	Dari Score (5)
Enrolled in Formal School			0.563***	0.684***	0.271***
			(0.089)	(0.094)	(0.086)
Head of Household's Child	0.053	-0.107	-0.133	-0.051	-0.235*
	(0.073)	(0.140)	(0.145)	(0.151)	(0.141)
Female	-0.208***	-0.688***	-0.572***	-0.572***	-0.464***
	(0.032)	(0.062)	(0.062)	(0.062)	(0.068)
Age	0.046***	0.287***	0.262***	0.235***	0.254***
	(0.010)	(0.019)	(0.019)	(0.020)	(0.018)
Duration of Family in Village	-0.001	-0.005**	-0.004**	-0.003	-0.006***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Family Identifies as Farsi	-0.031	0.085	0.099	0.126	0.039
	(0.047)	(0.084)	(0.084)	(0.082)	(0.095)
Family Identifies as Tajik	0.013	0.089	0.08	0.091	0.048
	(0.046)	(0.086)	(0.083)	(0.083)	(0.092)
Family Farms	-0.057	-0.035	-0.006	-0.007	-0.003
	(0.048)	(0.083)	(0.084)	(0.088)	(0.084)
Age of Household Head	-0.004**	-0.004	-0.002	-0.004	0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Years of Ed of Household Head	0.001	0.039***	0.039***	0.035***	0.037***
	(0.006)	(0.010)	(0.010)	(0.010)	(0.011)
Number of People in Household	0.003	-0.011	-0.013	-0.009	-0.017
	(0.008)	(0.015)	(0.014)	(0.013)	(0.016)
Jeribs of Land Owned by Household	0.020*	0.062***	0.049**	0.029	0.069***
	(0.011)	(0.021)	(0.020)	(0.022)	(0.021)
Number of Sheep	0.008**	0.014***	0.011**	0.010*	0.011**
	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)
Distance to Nearest Formal School	-0.048***	-0.075***	-0.050*	-0.048*	-0.042
(Non-NGO School)	(0.017)	(0.029)	(0.029)	(0.029)	(0.032)
Constant	0.166	-1.715***	-1.788***	-1.630***	-1.699***
	(0.151)	(0.278)	(0.282)	(0.294)	(0.278)
Observations	708	653	653	653	653
R-squared	0.17	0.41	0.46	0.45	0.33

Note: This table contains the estimated correlations between enrollment in formal school, test scores, and demographic characteristics in the control villages. All coefficients are estimated by regressing the indicated dependent variable on the listed demographic characteristics using an OLS regression with standard errors clustered at the household level. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 3: Demographic Characteristics by Research Group

	All Children			Only Children Tested		
	Treatment Average	Control Average	Estimated Difference	Treatment Average	Control Average	Estimated Difference
Panel A: Child Level Variables						
Head of Household's Child	0.935	0.911	0.024	0.939	0.917	0.022
			(0.019)			(0.019)
Female	0.474	0.455	0.02	0.495	0.475	0.02
			(0.025)			(0.026)
Age	8.321	8.312	0.009	8.323	8.303	0.02
			(0.068)			(0.072)
Panel B: Household Level Variables						
Duration of Family in Village	30.302	27.594	2.709**	30.239	27.852	2.387*
			(1.235)			(1.262)
Family Identifies as Farsi	0.208	0.209	-0.001	0.209	0.202	0.007
			(0.032)			(0.033)
Family Identifies as Tajik	0.243	0.208	0.035	0.245	0.214	0.031
			(0.033)			(0.034)
Family Farms	0.717	0.727	-0.01	0.709	0.721	-0.013
			(0.036)			(0.038)
Age of Household Head	40.142	39.97	0.172	40.268	39.839	0.428
			(0.887)			(0.910)
Years of Ed of Household Head	3.315	3.076	0.239	3.296	3.085	0.211
			(0.290)			(0.303)
Number of People in Household	8.399	7.818	0.581***	8.462	7.779	0.682***
			(0.225)			(0.231)
Jeribs of Land Owned by Household	1.345	1.274	0.071	1.345	1.239	0.106
			(0.128)			(0.128)
Number of Sheep	7.552	5.631	1.921***	7.408	5.755	1.653***
			(0.603)			(0.612)
Distance to Nearest Formal School	2.91	3.163	-0.253***	2.923	3.161	-0.238***
(Non-Community-Based School)			(0.085)			(0.088)

Note: This table contains average demographic characteristics divided by research group. The first three columns includes all children in the sample while the second three columns include only those children that were tested as part of the surveying process. All differences are estimated using equation (1) with standard errors clustered at the household level. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 4: Attrition Patterns by Research Group

	Attritors less Non-Attritors			Non-Attritors		
	Treatment Difference	Control Difference	Difference in Difference	Treatment Average	Control Average	Estimated Difference
Panel A: Attrition Rates						
	0.174	0.162	0.011			
	(0.014)	(0.014)	(0.022)			
Panel B: Child Characteristics						
Head of Household's Child	-0.001	-0.049	0.048	0.935	0.919	0.016
	(0.023)	(0.029)	(0.045)			(0.020)
Female	-0.04	-0.024	-0.016	0.481	0.459	0.023
	(0.047)	(0.051)	(0.069)			(0.027)
Age	0.528	0.229	0.299	8.229	8.275	-0.046
	(0.155)	(0.167)	(0.235)			(0.076)
Panel B: Household Characteristics						
Duration of Family in Village	-5.302	-2.671	-2.63	31.224	28.028	3.197**
	(1.452)	(1.593)	(2.661)			(1.295)
Family Identifies as Farsi	-0.003	0.031	-0.034	0.209	0.204	0.005
	(0.038)	(0.041)	(0.067)			(0.034)
Family Identifies as Tajik	-0.054	-0.051	-0.003	0.252	0.216	0.036
	(0.040)	(0.041)	(0.064)			(0.036)
Family Farms	-0.032	0.035	-0.067	0.723	0.722	0.001
	(0.043)	(0.045)	(0.074)			(0.038)
Age of Household Head	-1.382	1.105	-2.487	40.382	39.791	0.591
	(1.055)	(1.162)	(1.841)			(0.937)
Years of Ed of Household Head	-0.372	-0.054	-0.318	3.379	3.084	0.295
	(0.333)	(0.358)	(0.565)			(0.309)
Number of People in Household	-0.563	-0.27	-0.293	8.497	7.862	0.635***
	(0.275)	(0.261)	(0.460)			(0.245)
Jeribs of Land Owned by Household	0.259	0.062	0.197	1.3	1.264	0.036
	(0.147)	(0.166)	(0.261)			(0.138)
Number of Sheep	-0.268	-1.709	1.441	7.599	5.909	1.690**
	(0.763)	(0.710)	(1.116)			(0.655)
Distance to Nearest Formal School	-0.258	0.161	-0.418**	2.955	3.137	-0.182**
(Non-Community Based School School)	(0.107)	(0.111)	(0.180)			(0.087)

Note: This table contains average demographic characteristics divided by research group of attriting and non-attriting students. The first three columns compare the average characteristics of attriting and non-attriting students with the difference in attrition patterns estimated using equation (3) with standard errors clustered at the household level. The second three columns report the average characteristics of non-attriting students. The differences in column 6 are estimated using equation (1) with standard errors clustered at the household level. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 5: Enrollment in NGO and Formal Schools

	Treatment Average	Control Average	Difference	Difference w/ Controls
Panel A: All Children				
Community-Based School, Fall 2007	0.564	0	0.564***	0.556***
			(0.020)	(0.021)
Formal School, Fall 2007	0.735	0.268	0.467***	0.421***
			(0.028)	(0.029)
Formal School, Winter 2008	0.416	0.007	0.410***	0.420***
			(0.025)	(0.024)
Panel B: Tested Children				
Community-Based School, Fall 2007	0.581	0	0.581***	0.574***
			(0.021)	(0.021)
Formal School, Fall 2007	0.739	0.27	0.470***	0.423***
			(0.028)	(0.030)
Formal School, Winter 2008	0.432	0.007	0.425***	0.437***
			(0.026)	(0.025)

Note: This table contains estimates of the effect of the treatment on students' enrollment in NGO schools and formal schools. The first two columns contain the average enrollment rates in the treatment and control groups. The second two columns contain the estimated differences using equation (1) and equations (2) respectively clustering at the household level. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 6: Enrollment in Individual Institutions

	Treatment Average	Control Average	Difference	Difference w/ Controls
Community-Based School, Fall 2007	0.564	0	0.564***	0.556***
			(0.020)	(0.021)
Government School, Fall 2007	0.161	0.267	-0.106***	-0.146***
			(0.025)	(0.026)
Mosque School, Fall 2007	0.051	0.222	-0.171***	-0.165***
			(0.022)	(0.023)
Mosque School, Winter 2007/8	0.788	0.863	-0.075***	-0.079***
			(0.024)	(0.024)
Madrassa, Fall 2007	0.01	0.001	0.009**	0.011**
			(0.004)	(0.005)

Note: This table contains estimates of the effect of the treatment on students' enrollment in individual educational institutions. The first two columns contain the average enrollment rates in the treatment and controls groups. The second two columns contain the estimates differences using equation (1) and equation (2) respectively clustering at the household level. Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 7: Effects of Treatments on Test Scores

	Treatment Average	Control Average	Difference	Difference w/ Controls	TOT w/ Controls
Panel A: Fall 2007 Survey					
Total Score	0.58	-0.007	0.587***	0.506***	1.194***
			(0.059)	(0.054)	(0.127)
Math Score	0.614	-0.005	0.620***	0.549***	1.296***
			(0.058)	(0.055)	(0.126)
Dari Score	0.418	-0.008	0.426***	0.344***	0.813***
			(0.058)	(0.055)	(0.131)
Panel B: Spring 2008 Survey					
Total Score	0.598	0.003	0.596***	0.528***	1.186***
			(0.060)	(0.056)	(0.125)
Math Score	0.671	0.004	0.667***	0.611***	1.373***
`			(0.060)	(0.056)	(0.128)
Dari Score	0.456	0.001	0.455***	0.379***	0.852***
			(0.061)	(0.056)	(0.125)

Note: This table contains estimates of the effect of the treatment on children's test scores. The first two columns contain the average enrollment rates in the treatment and controls groups. The second two columns contain the estimates differences using equation (1) and equation (2) respectively clustering at the household level. The final estimated intent to treat effect of enrollment in a formal school is estimated through equations (4) and (5). Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 8: Effect of Distance on Enrollment and Test Scores

Dependent Variable	(1) **Distance to School**	(2) **Formal Enrollment**	(3) **Formal Enrollment**	(4) **Formal Enrollment**	(5) **Total Score**
Treatment	-2.697***				
	(41.800)				
Distance to Nearest Formal School		-0.129***	-0.159***	-0.160***	-0.191***
		(14.940)	(14.600)	(14.140)	(9.530)
Head of Household's Child	0.019	-0.001	-0.009	-0.015	-0.021
	(0.200)	(0.020)	(0.190)	(0.280)	(0.210)
Female	-0.011	-0.116***	-0.117***	-0.122***	-0.557***
	(0.260)	(5.110)	(5.110)	(5.140)	(12.060)
Age	0.004	0.051***	0.051***	0.050***	0.308***
	(0.400)	(7.580)	(7.510)	(6.860)	(23.540)
Duration of Family in Village	-0.004*	-0.001	-0.001	-0.001	-0.004**
	(1.700)	(0.580)	(0.890)	(1.220)	(2.370)
Family Identifies as Farsi	-0.029	-0.038	-0.039	-0.044	-0.013
	(0.420)	(1.150)	(1.150)	(1.270)	(0.190)
Family Identifies as Tajik	-0.067	0.009	0.001	0.005	0.094
	(0.810)	(0.250)	(0.040)	(0.130)	(1.480)
Family Farms	0.229***	-0.024	-0.016	-0.012	-0.012
	(3.820)	(0.720)	(0.470)	(0.360)	(0.180)
Age of Household Head	0.002	-0.001	-0.001	-0.001	0.003
	(0.830)	(1.070)	(0.870)	(0.390)	(1.150)
Years of Ed of Household Head	0.017*	0.004	0.004	0.005	0.042***
	(1.840)	(1.000)	(1.030)	(1.130)	(5.320)
Number of People in Household	-0.002	0.001	0.001	0.001	0.002
	(0.180)	(0.240)	(0.140)	(0.210)	(0.150)
Jeribs of Land Owned by Household	0.008	< 0.001	< 0.001	< 0.001	0.015
	(0.390)	(0.030)	0.000	(0.050)	(0.930)
Number of Sheep	-0.005	0.003	0.003	0.003	0.010**
	(1.170)	(1.460)	(1.270)	(1.400)	(2.310)
Constant	2.930***	0.316***	0.388***	0.389***	-1.988***
	(14.670)	(3.180)	(3.800)	(3.550)	(9.790)
Observations	1490	1490	1490	1374	1374
R-squared	0.75	0.27	0.27	0.27	0.39
Model	OLS	OLS	IV	IV	IV

Note: This table contains the estimated effect of distance on enrollment and test scores. Column one contains the first stage estimate of the relationship between receipt of the treatment and distance to the nearest formal school (equation (5)). Column two contains the OLS estimate of equation 5. Column three contains the instrumental variables estimate of the effect of distance on enrollment using equations (4) and (5). Column four contains the same estimate as column three but only for students taking the fall 2007 exam. Finally, column five estimates the effect of distance on children's test scores using equations (4) and (5). Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.

Table 9: Treatment Effects by Gender

Dependent Variable	(1) CBS Enrollment	(2) Formal Enrollment	(3) Total Score	(4) Total Score	(5) Formal Enrollment	(6) Total Score
Treatment	0.464***	0.349***	0.384***			
	-16.35	-9.05	-5.28			
Treatment * Female	0.226***	0.153***	0.249***			
	-6.27	-3.37	-2.73			
Formal School				1.112***		
				-5.93		
Formal School * Female				0.142		
				-0.74		
Distance to Nearest Formal School					-0.132***	-0.148***
(Non-NGO School)					-9.9	-5.59
Distance * Female					-0.058***	-0.088***
					-3.48	-2.67
Head of Household's Child	-0.016	-0.014	-0.02	-0.003	-0.011	-0.023
	-0.41	-0.26	-0.2	-0.03	-0.22	-0.23
Female	0.001	-0.200***	-0.685***	-0.485***	-0.024	-0.414***
	-0.19	-6.14	-10.83	-4.32	-0.69	-5.61
Age	0.002	0.049***	0.307***	0.250***	0.052***	0.308***
	-0.29	-6.94	-23.78	-16.29	-7.57	-23.73
Constant	-0.011	0.048	-2.381***	-2.413***	0.336***	-2.066***
	-0.14	-0.45	-11.76	-11.62	-3.24	-10.03
Household Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1374	1374	1374	1374	1490	1374
R-squared	0.43	0.3	0.4	0.47	0.27	0.39
Model	OLS	OLS	OLS	IV	IV	IV

Note: This table contains estimates of the effect of the program by gender. Columns one, two and three show the effect of the treatment on community-based school enrollment, formal school enrollment, and test scores using equation (2). Column four contains the effect of the treatment on the treated children using equations (4) and (5). Finally, columns five and six contain estimates of the effect of distance on enrollment and test scores also using equations (4) and (5). Statistical significance at the one, five, and ten percent levels is indicated by ***, **, and * respectively.