Parental Human Capital and Effective School Management

Evidence from The Gambia[*]

Moussa P. Blimpo[1]   David K. Evans[2]   Nathalie Lahire[2]

April 2015

Abstract: Education systems in developing countries are often centrally managed in a top-down structure. In environments where schools have different needs and where localized information plays an important role, empowerment of the local community may be attractive, but low levels of human capital at the local level may offset gains from local information. This research evaluates the effectiveness of a program that provided a grant and comprehensive school management training to principals, teachers, and community representatives in a set of schools. In order to separate the effect of the training from the grant, a second set of schools received the grant only with no training. A third set of schools served as a control group and received neither intervention. Each of 273 Gambian primary schools were randomized to one of the three groups. Three to four years into the program, the full intervention led to a 21% reduction in student absenteeism and a 23% reduction in teacher absenteeism, but produced no impact on student test scores. The effect of the full program on learning outcomes is strongly mediated by baseline local capacity, as measured by adult literacy. This result suggests that, in villages with high literacy, the program may yield gains on students' learning outcomes. Receiving the grant alone had no impact on either test scores or participation.

Keywords: Education, Management, School-based management
JEL Classification: O15, I21, C93

[1] University of Oklahoma. moussa.blimpo@ou.edu
[2] World Bank. devans2@worldbank.org and nlahire@worldbank.org

1. Introduction

Every year, billions of dollars are spent to provide services to the poor in low-income countries. Unfortunately, there is a long-standing record of failures in delivery systems, whether in education, health, or other sectors. Empowerment of local communities in school management has received growing attention from both academics and practitioners in developing countries as part of a broad and global program to improve service delivery to the poor by involving them directly in the delivery process (World Bank 2004). The quality of local school management has been shown to be strongly associated with favorable education outcomes across countries (Bloom et al. 2014). In Africa, countries including Ghana, Niger, Senegal, Madagascar, Kenya, Rwanda, and Mozambique have already embraced variants of school autonomy in their education systems (Bruns, Filmer & Patrinos 2011).

In this research, we assess the medium-run impact of a program seeking to empower local communities in school management in the Gambia. On the one hand, local leadership may have significant additional information relative to the central authorities about local needs, local politics, and other constraints. Local management also may increase accountability (Bruns, Filmer & Patrinos 2011), as was observed and demonstrated with a school-based management and accountability program in Mexico (Gertler, Patrinos, & Rubio-Codina 2012). These would suggest that the program may be effective in improving student learning. On the other hand, local leadership or members of the community may also lack the competency (relative to central leadership) to design or implement the processes necessary to tackle local problems, suggesting that the program could be ineffective. The net effect of such a policy is ambiguous.[3] This paper uses a large field experiment in the Gambia to evaluate and draw lessons from a comprehensive school management and capacity building program – called Whole School Development (WSD). The intervention and subsequent data collection were carried out between 2007 and 2011.

In WSD schools, principals, certain teachers, and members of the communities received comprehensive training in school management. During this training, the schools' stakeholders (including the community) developed school management plans addressing short- and long-term goals in each of these areas. A national semi-autonomous WSD unit associated with the Ministry of Education guided them. In order to help schools initiate the implementation of their plans, the Ministry of Education provided a grant worth approximately US$500. To separate the effect of the grant from that of the training, another set of schools received a grant of the same size without the accompanying training component (called Grant-only schools).

---

[3] Retrospective evaluations of such complex programs present many challenges, but early evidence from El Salvador's community-managed schools program found positive impacts on participation and language skills (Jimenez and Sawada 1999).

In addition, a new school constitution had been developed by the Ministry of Education as part of its new School Management Manual (SMM) to enhance cooperation between teachers and the community. Acceptance of the new constitution was a prerequisite for receipt of the grant. All schools receiving grants (both schools with WSD plus grant and Grant-only schools) were directed to use the grant towards some aspect of school development that related directly to teaching and learning (i.e., constructing teacher housing would not be an acceptable use). Finally, the control schools received neither a grant nor the management training. We randomly assigned each of 273 Gambian basic cycle schools to one of the three groups.

At the end of the 2011 school year, three to four years into the program, we found no effect of the WSD intervention on learning outcomes, measured by scores on a comprehensive test in Mathematics and English. However, we found that the intervention did lead to a reduction in student absenteeism of nearly 5 percentage points from a base of 24 percent, and a reduction in teacher absenteeism of about 3 percentage points from a base of about 13 percent. We found no effect of the Grant-only intervention, relative to the control, on test scores or participation. If the reduction in student absenteeism in the WSD schools led to increased attendance of students with poorer performance, then the average treatment effect on test scores would be biased downward. To correct for this potential selection bias, we used Lee's (2009) trimming procedure to calculate the upper and lower bounds of the treatment effect on test scores. Our estimates indicate that, once corrected for selection, the average treatment effect ranges from -0.19 to 0.17 standard deviations for Mathematics and -0.16 to 0.26 standard deviations for English. Given that the bounds are roughly centered on zero, we take zero as our preferred and conservative point estimate.

We analyzed the importance of baseline local capacity in mediating the effect of the WSD. As mentioned earlier, theory would predict that, all else equal, the WSD is more effective in areas with higher baseline capabilities. We interacted the intervention dummies with average district level adult literacy in 2006. The estimates yield a positive and significant effect of the interaction term. The finding is qualitatively the same when we replace district level adult literacy by the share of School Management Committee (SMC) members with no formal education (i.e., cannot read or write): In that case, we find a negative and significant effect of the interaction term. Our findings suggest that the WSD can work in areas with higher adult literacy at baseline. Our point estimates suggest that a minimum of 45% adult literacy is needed for the WSD to begin showing effects on learning outcomes. We found no interaction effect for the Grant-only intervention. In summary, we find little to no evidence that a comprehensive intervention such as the WSD can improve learning outcomes, except when baseline capacity is sufficiently high.

This paper adds to the literature on interventions to increase community involvement in schools. The findings are consistent with Banerjee et al. (2010) who compare three interventions that aim to increase community involvement in the Indian context, where the central government is expanding the number of schools that are organized locally.

They found these interventions to have no effect on beneficiaries' participation or on learning outcomes.

In contrast, a recent study in Kenya compared different interventions involving additional resources, teacher incentives, and some level of institutional changes (Duflo et. al. 2014). They found that training the community to specifically monitor teachers, combined with reduced class size and teacher incentives, yielded significant gains in various education outcomes. They also found that hiring additional teachers reduced the effort of existing teachers. However, where communities were involved in monitoring, the negative impact on teachers' effort dropped significantly, leading to improvement in learning outcomes.

Our findings also contrast with those of Bjorkman and Svensson (2009) who evaluated another intervention to enhance community engagement in the health sector in Uganda. They provided report cards (on health care providers) to members of treatment communities and encouraged them to define monitoring strategies. One year into the program, they found large effects on health outcomes. Why do some of these – apparently similar – interventions seem to work whereas other – such as the WSD – did not? Beside the specificities of the contexts and the interventions, there is at least one fundamental difference between these two sets of interventions: the extent to which the intervention is simple and focused on one or a few specific areas. Whereas the WSD is a comprehensive (and relatively complex) program, these two interventions, and many similar interventions that worked, are focused on one main dimension: monitoring.

There are other potential reasons why the WSD did not work to improve learning outcomes on average. First, in low-income countries such as the Gambia, other inputs that enter the educational production function such as teacher quality and content knowledge might be low and thus constitute binding constraints that prevent other policies from functioning well. For example, in the course of this evaluation, Gambian teachers agreed to take a sixth-grade level content knowledge test and revealed overall poor outcomes. In addition, due to resource constraints, a large number of schools function in double shifts and the total instructional time is less than 80% of what is recommended.

Second, in low-income countries, the problem of local capture has often been pointed out in the literature as one of the main drawbacks of decentralization (Bardhan and Mookherjee 2002; Gugerty and Kremer 2008; Reinikka and Svensson 2004). However, we find no evidence of this issue in the context of the Gambia when we analyze the school finances and the disbursement process. The WSD program put in place a mechanism to prevent the misuse and misappropriation of school funds. All expenses were required to be approved by the SMC and the regional directorate. Schools were required to subsequently submit the receipts to the regional directorate. In addition, there were officials at the regional directorate, called "cluster monitors," whose role was to monitor activities at the school level and report back to the director. There is no evidence suggesting that political economy forces, such as local capture, were at play.

Finally, even in an environment where local capture is limited or controlled, local capacity to make informed decisions and effectively implement them is crucial to the success of decentralization policies. In high-income countries such as the United States, conventional wisdom suggests that institutional arrangements that favor and foster accountability, competition, and autonomy are the most effective in improving schools (Hanushek and Woessman, 2007 & 2009). Differences between the high and low-income countries, and even between India and the Gambia, render extrapolation from existing evidence to poor country settings difficult. The interaction effects reported earlier suggest that baseline local capacity may constrain the benefits from local empowerment. We conclude that a combination of low baseline local capability, the complexity of the intervention, and the low quality of other educational inputs are the main factors explaining the limited impact of the intervention. School-based management models will need to be appropriately adapted to the needs of local communities.

2. The context

This section combines administrative data with our baseline data to describe the education system in The Gambia. Basic education in The Gambia lasts nine years. The first six years are called Lower Basic and the following three years are Upper Basic. Upon completion of basic education, students take a national exam (9th grade exam) that determines admission to high school. High school lasts an additional three years.

The education sector in The Gambia has been growing rapidly in recent years. The total number of students enrolled in the formal education system doubled between 1998 and 2010. Nearly every community has its own lower basic school or has one within a five-kilometer radius. The basic infrastructure (classrooms, tables, chairs, water) is in general sufficient even in rural areas. However, due to the increased enrollment, many schools have adopted a double shift system where one group of students comes in the morning and the other group in the afternoon.

In terms of organization, there is a Ministry of Basic and Secondary Education (MoBSE) in charge of the education system up to 12th grade. The country is organized in six administrative regions: five regions outside the capital plus the district of Banjul (the capital city). Each of the regions has a regional educational office with a regional director. The regional directors are the key liaisons between the schools in their region and the ministry. They ensure the monitoring of activities at the school level and collect key indicators on a regular basis.

The baseline data from this research (gathered in 2008) include specific information about Gambian schools (Adebimpe, Blimpo, and Evans, 2009). Those data demonstrate that overall the basic infrastructure of schools was in good condition.[4] The main buildings

---

[4] These assessments are based on visual observation by the enumerators. We limited self-reported information whenever possible. For example, when inquiring about management

5

(classrooms and staff headquarters) were overall in good condition throughout the country. Of the 273 schools visited, 9% required some minor repairs for the walls, roofs, floors etc. One percent of the schools was in very bad condition and needed total rehabilitation; these schools were all located in one region. In another region, 15% of the schools had buildings that needed minor repairs. In 97% of the 526 classrooms visited, most of the students were seated on a chair with a table. The teaching areas were equipped with a chair and a table in 92% of the classrooms visited. The student-teacher ratios were similar across regions at about 40 students per teacher.

At the baseline survey, we looked at recordkeeping as one proxy for management. When the head teacher was the respondent, 69% reported keeping financial records and were able to show them. In the absence of the head-teacher, we interviewed the deputy head teachers. In those cases (i.e., when the head teacher was absent), only 30% of them reported that the school kept records of finances and were able to show them. Forty-one percent of schools conducted classroom observation to ensure the quality of the teaching and were able to show records that confirmed it. All the schools reported the existence of some form of Parent-Teacher Association (PTA); however, 65% of PTAs have no funding. Head teachers were asked to report the most important challenge that the school faced in its effort to provide proper education to the student. The most frequent responses were the lack of resources (34%) and the lack of proper teacher training (14%).

Absenteeism is high for both students and teachers but is comparable to other low-income countries. Within the surveyed schools, teacher absenteeism ranged from about 12% of teachers absent on the day of the survey in two regions to about 30% in another region. In addition, during the classroom visits, 32% of the teachers reported having missed at least one day of class during the previous week. Forty-eight percent of teachers had a written lesson plan. Student absenteeism is measured as the percentage of the class that was absent on the day of the survey in two randomly selected classes in each school: specifically, a randomly selected classroom of classes 4 and 6 where possible; where not possible, a randomly selected other class. In the 526 classroom visits, student absenteeism ranged from about 20% of the total number of students enrolled in some regions to nearly 40% in another.

Learning assessments have revealed poor learning outcomes: For example, the 2007 Early Grade Reading Assessment found that almost 50% of third graders could not correctly read a single word (USAID et al. 2008). Hence there is strong demand to improve learning outcomes. Within this study, in terms of both literacy and numeracy, student performance is lower than expected (per the curriculum) in Grade 3 but improves substantially by Grade 5, indicating that – at least – students are learning in school. There was considerable heterogeneity in student performance within each grade, particularly in math skills. In almost all tests, girls under-performed boys by about 3 percentage points.

---

practices such as good recordkeeping, in addition to yes or no answers, enumerators recorded a third option that consisted of visually confirming the existence of the relevant records.

On average, third graders are 10 years old and the fifth graders are 12 years old. Half of the students live in homes with improved latrines. Only 20% of the students reported having electricity. Ninety percent of students had a radio at home, 83% of households owned a telephone,[5] and 69% owned a bicycle.

3. Experimental design

3.1. The intervention arms

The main intervention evaluated in this paper is a holistic school management capacity building program called Whole School Development (WSD).[6] This intervention consists of the distribution of management manuals, a comprehensive training component, and a grant to help implement the activities in the first year. In order to be able to separate the impact of the capacity building component from the grant, a second intervention group received the grant but did not receive training. We compare these two interventions to a control group that received neither the grant nor the training. Table 1 provides a summary of the key elements of the intervention arms, and Table 2 summarizes the project timeline.

3.1.1. The Management manual

The school management manual (SMM) is a comprehensive guide to management practices both within the school and for interactions with other stakeholders at the community, regional, and national levels. International experts developed the manual together with national officials and stakeholders at the local level, including teachers. The manual addresses six specific topics pertaining to the management and functioning of schools: school leadership and management, community participation, curriculum management, teacher professional development, teaching and learning resources (e.g., textbooks and libraries), and the school environment. All these aspects are integrated in a three-step cycle for effective school management. The first step is information gathering and analysis. This step provides information as to what kind of data and information should be collected by schools on a regular basis (e.g., monitoring learning outcomes and absenteeism). It emphasizes how to analyze the data and then to create a plan for short-term and long-term solutions to school problems. The second step is the implementation of the resulting plan. The third step involves effective monitoring of the plan that is being implemented and adjustments along the way. The SMM advocates for strong, broad inclusiveness in school decision making. The manual was provided to all schools participating in the study.

3.1.2. The Management Training

---

[5] Either the household had a landline or a person in the household possessed a mobile phone.
[6] The WSD intervention has previously been implemented in South Africa (Bayona & Sadiki 1999).

The management training and capacity building are the centerpiece of the WSD intervention. The principals, teachers, and representatives of parents and students received training in six areas of school management, also described in the school management manual. The six areas were (1) community participation, (2) learner's welfare and school environment, (3) curriculum management, (4) teaching and learning resources, (5) teachers' professional development, and (6) leadership and management. In the course of this training, participants developed a local school development plan addressing various areas with guidance from the trainers and the supervision of the WSD unit within the Ministry of Education.

The training used a cascade method. First, the experts who developed the SMM trained twenty people ("master trainers") at the national level. Second, the master trainers conducted regional trainings of "cluster monitors" (school inspectors over a cluster of schools), school directors, and some senior teachers. Then those regional trainees carried out a local training with the school management committee, senior teachers, and – in some cases – a student representative. The training lasted between 10 and 20 days, with sessions split across several periods. The initial sessions of the local trainings were supervised by experts who had developed the SMM. Since most parents do not speak, read, or write English, the training put emphasis on local languages and drawings (See Figure 1) to convey the messages more effectively.

3.1.3. The Grant

Some of the activities suggested in the manual and included in the school development plans, like workshops, might require financial resources. Over time, the funding for these activities was expected to come from the school budget and locally raised funds. However, during the first year, the intervention schools were provided with a grant to serve as a catalyst for school improvement. A grant of US$500 was given to all the schools in the WSD and the Grant-only groups after a school development plan was presented. The schools were required to spend the funds on activities pertaining broadly to learning and teaching. The schools informed the regional office about their spending plans and submitted the receipts. This grant represents about 16 months worth of salary for a first grade teacher without experience or about 14.5 months worth of salary of a first grade teacher with five years of experience. It represents less than 5% of the average annual school budget.

3.2. Sampling

The sample in this study is the census of lower basic public and government-aided schools in regions 2, 3, 4, and 6 (276 schools) in the Gambia (Figure 2). The two regions that were excluded from the study were Region 1, which is essentially only the capital city and was excluded on the basis that it was too urban and distinct from the rest of the country, and

Region 5, because it was used extensively to pilot the WSD prior to the large randomized experiment. Of the 276 schools, one school was excluded from the sample because it was very small and had only a few students in grades 1 and 2. Another school was closed but still appeared on the official list of schools.

Figure **3** summarizes the sampling procedure.

Of the 273 remaining schools, 90 schools were randomly assigned to the WSD treatment, 94 schools to the Grant-only treatment, and 89 schools served as the control group. The schools were clustered in groups of 2 or 3 schools on the basis of geographic proximity to limit contamination while allowing useful exchange and cooperation between nearby schools.[7] Because this represents the universe of schools meeting the inclusion criteria, rather than a sample, clustering of groups of schools is unnecessary in the subsequent analysis.[8] The randomization was further stratified by school size and accessibility.[9] Each group proved to be similar at baseline, as discussed in detail in Section 5.1. As all schools remained in the study between baseline and endline, there is zero attrition.

4. Data

The Gambia Bureau of Statistics, under the supervision of the research team, collected the data for this study. The baseline data were collected in 2008 at the onset of the study, the first round of follow-up data were collected in 2009, the second round of follow-up data were collected in 2010, and the end-line data were collected in 2011 (Table 2).

In the 2009 follow-up, data were collected in the WSD and Control schools only. The Grant-only schools were not visited at that time because grant disbursement was delayed in one region during the first year, and many schools that had received their grant had not yet used it.[10] This problem of slow disbursement of education grants by local committees was also observed in Kenya (Conn et al. 2008).

At each round, teams of enumerators arrived unannounced (in order to avoid strategic attendance by teachers and students) at each school and collected information about the

---

[7] At the regional level, schools that are close to one another are assigned a "cluster monitor" who serves as a liaison between the regional directorate and those schools. The cluster monitor is encouraged to promote good practices among the schools she is assigned to.

[8] Furthermore, the intracluster correlation for test scores and absenteeism is much higher (55-80% higher) at the school level than at the level of school clusters.

[9] The Ministry defines accessibility through "hardship status". Schools that are most remote receive an allowance from the Government, as discussed in Pugatch & Schroeder (2014).

[10] This information was obtained from the regional directorates who were the key intermediaries for the grant disbursement process.

school and the students, conducted classroom observation, and gave a literacy and numeracy test.[11] Unless otherwise indicated, the following data were collected at each of the four rounds of data collection. Table 3 summarizes the data collected in each wave.

4.1. School data

The data on the school as a whole were obtained through enumerator observation and a comprehensive interview with the head teacher or – in the absence of the head teacher – the teacher in charge of the school at the time. The directly observed information includes the condition of the buildings, the number of classrooms and other facilities, etc. Information from the head teacher included school finances, record keeping, community participation, management practices, etc. To improve the accuracy of the information collected, we requested to see written records to substantiate responses whenever applicable.

4.2. Classroom visits

In each school, we randomly selected two classrooms for observation. The goal of the classroom visit was to gather information about teaching practices, the classroom environment, and student participation. It also served to substantiate the absenteeism data from the administrative records by comparing the student register to the number of students present in the classroom. Each classroom visit lasted fifteen minutes, followed by a five-minute interview with the teacher.

4.3. Student written literacy and numeracy test

Forty students were selected randomly at each school and were given a written numeracy and literacy test. At the baseline, we tested twenty third-grade students and twenty fifth-grade students at each school. Third- and fifth-graders were selected as these are the earliest grades regularly evaluated by the Gambian government. At the first follow up in 2009, we gave the test to students in fourth and sixth grades to allow for tracking of the baseline students. At the second follow-up in 2010, the test was given again to third and fifth grade students because much of the original cohort would have completed primary school. In total, 8,959 students were tested at baseline, roughly evenly distributed across the three treatment groups.

4.4. Student interview and oral literacy test

Of the forty students who took the written test, ten were randomly selected to take an orally administered reading and comprehension test and to participate in an interview about their socio-demographic characteristics, school performance, and other information. These students were tracked in 2009 in the WSD and Control schools, and in

---

[11] The schools were given a range of time during which a team of enumerators would visit them. The actual dates were not disclosed.

2010 in all the schools whenever possible.[12] Students for the pupil interview were selected randomly from among those who participated in the written test. At baseline, 2,696 students were interviewed in total: 879 from WSD schools, 920 from Grant-only, and 897 from the control schools.

4.5. Teacher content knowledge

In 2009, we tested teacher knowledge of content: The test was similar to the students' written test, with additional questions drawn from Gambian secondary school reading and math textbooks. A short background interview was also administered to the teachers who took the test.

4.6. Qualitative data

In 2010, we added many open-ended questions to the head teacher interviews to collect some information about their views regarding school management. We addressed similar questions to parents or caregivers in a few households whose children were in the relevant schools. The research team was also heavily involved on the ground for the entire first year of this program; the associated conversations with the government, the schools, and the communities add important information that is useful for a better understanding of the findings.

5. Identification, empirical strategy, intermediate outcomes

5.1. Identification and group comparison

In the design of a field experiment, the goal of employing random assignment to allocate participation in the program is to achieve a situation in which each of the groups has similar characteristics – both observed and unobserved – before the implementation of the program. If the treatment and control groups are balanced at baseline, then differences in teaching activities and student learning outcomes between the groups in the follow up survey can be attributed to the WSD and Grant-only programs, rather than to some pre-existing difference between the groups. Using the data from the baseline survey, we examine observed characteristics across the different groups.

We first compare the outcome variable at baseline across groups. Figure 4 shows the distribution of test scores of fifth-grade students on a written test in English, Math, and a combined score. It shows that the baseline performance level of student, across groups, comes from the same distribution. The t-test of comparison of means cannot reject the hypothesis that the underlying distribution of students' performance at the baseline has the same mean. Similarly, the Kolmogorov-Smirnov test of comparison of distribution does not reject the hypothesis that the distributions of students' performance are

---

[12] Most of the students in 5th grade at baseline had finished the basic cycle by the time of the second follow up.

identical across the three groups. We reach the same conclusion on the student reading outcomes. Fifth grade students were presented with a sixty-word text to read in one minute. Figure 5 shows the similarity of the distribution of reading outcomes across the groups. In addition to the students' baseline performance, we compare school and student characteristics across groups.

A list of indicators and their means across groups are included in Table 4 (school characteristics) and Table 5 (student characteristics). We observe no systematic differences across the groups. For example, the average size of the schools is comparable across groups and the average student-teacher ratio is nearly identical: There were 32 students per teacher in the WSD and Control schools versus 34 in the Grant-only schools. Out of 17 characteristics at the school level, the only significant difference is that the WSD schools on average reported 4.4 Parent-Teacher Association (PTA) meetings during the year prior to the survey versus 3.7 for both the Grant-only and the Control group. WSD communities held sensitization meetings about the program in advance of program implementation and the survey, which is the most likely explanation for this difference. There is no significant difference for cash and in-kind contributions across groups, which might be expected if the difference in meetings were an indication of greater baseline involvement more generally in WSD communities.

In terms of student characteristics, the groups are comparable as well. Third-grade students are a little over 10 years old and fifth-graders are about 12.5 years old in all three groups. The socioeconomic backgrounds of students, in terms of access to electricity at home, possession of a television, and access to a telephone are also comparable across groups. The percentage of students currently repeating a grade is identical (9%) in all three groups. We conclude that there are no apparent systematic differences across the treatment groups at the baseline. The random assignment to the different interventions groups means that there are also no expected systematic differences among the three groups in unobserved characteristics.

5.2. Main Empirical Strategy

Because of the random assignment of schools to the treatment groups, the following basic regression model provides the estimates of the causal effect of the interventions:

$$Outcome_{is} = \alpha + \beta_1 WSD_s + \beta_2 Grant_s + \varepsilon_{is} \qquad (1)$$

where $Outcome_{is}$ is the outcome of student i in school s, WSD_s = 1 if school s received the WSD intervention and 0 otherwise, and $GRANT_s$ = 1 if school s received the grant-only intervention and 0 otherwise. The error term ε_{is} is clustered at the school level to account for intra-school correlation of outcomes. The parameters of interest are β_1, which is the average effect of the WSD intervention on the outcome, and β_2, which is the average effect of the Grant-only intervention. A simple test of the null hypothesis – $H_0: \beta_1 = \beta_2$ – compares the WSD intervention to the Grant-only intervention.

5.3. Intermediate results

5.3.1. One year post-interventions

One year after the implementation of the WSD, we collected data in all the WSD and control schools. The goal of this round of data collection was to ensure that the WSD was properly implemented, to monitor the evolution of the process, and to collect some intermediate variables to assess the early impact. The key results described in this section are reported in Tables 6, 7, 8.

Most of the significant results at the school administration level are focused around take-up of the WSD program in the WSD schools. We assessed take-up by looking at basic elements that indicate whether the WSD program is functioning or not.[13] There is a higher rate of establishment of various school management committees (SMC) in WSD schools, as recommended by the School Management Manual (Table 6). For example, 84% of the WSD schools had set up a curriculum management committee whereas only 51% of the control schools did so. (The committees in the control group are often different in nature and reflect the school organization in place prior to this research.) Similarly, for each of the other SMCs, we observed statistically significant differences in favor of the WSD. Only about one-third of the schools in each group had adopted and actually implemented the new PTA constitution, with a 3-percentage point edge in the WSD schools.

In terms of intermediate outcomes, the control schools appeared to perform better in teacher preparedness one year into the program (Table 7). We observed teachers' written lesson notes for the day of the visit in more control classrooms (41%) than in the WSD classrooms (32%). We also observed 11% more lesson plans in the control classrooms than the WSD classrooms. Both of these results are significant.[14] It could be that new committee work associated with the set-up of the WSD program actually took teachers away from classroom preparation. (This is consistent with the fact that significant differences on teacher preparation disappear in subsequent observations, longer after the program was established.)

Absenteeism remained pervasive (Table 7). About 25% of the students were missing, when we compared the number of students present to the number of students listed on the register. We also picked five days randomly from the register and found an average of nearly 38% recorded absenteeism over those 5 days, nearly identical in both groups. More teachers in the control group (7% more) reported having missed at least one day of class in the previous week. Teacher absenteeism remained the same as at the baseline in the control group (32% of teachers reported having missed a day during the previous

---

[13] The control schools were given the basic manual of the WSD, but that they did not receive the training and the grant.

[14] In this context, the "lesson plan" is the weekly or monthly outline of topics to be taught, whereas the "lesson note" is the document outlining the specific activities for a given day.

week) whereas it dropped by 6 percentage points in the WSD group, according to teacher reports. However, the average percent of teachers absent over 5 random days, based on school records, indicates relatively low absenteeism (6%) and no difference between across groups (Table 7).

We found no difference between the two groups in terms of student performance (

Table 8). Fourth graders read about 24 words per minute and sixth graders read 41 words. Research suggests that about 45 to 60 words per minute are required for comprehension (Abadzi 2008).

These findings show – unsurprisingly – a higher rate of adoption of the school organization recommended by WSD in the WSD schools and its components within the WSD group compared to the control group. No differences were observed regarding student performance, although it would likely be too early to observe such an effect at that point. At the very least, this indicates that the program was implemented as planned.

5.3.2. Two years post-interventions

In this section, we present the impact of the intervention on student learning outcomes, teaching practices at the school level, and school management two years into the interventions in all three groups.

The estimates of the average treatment effect (Table 9) indicate that neither the WSD nor the Grant-only interventions had any impact on student learning outcomes two years after their implementation. Student performance in all groups remains relatively poor and comparable to baseline levels. This is also true for the control group, which rules out the possibility that the control group may have improved along with the treatment groups over the two years but due to reasons other than the intervention.

Even though we observe no average treatment effect, it is possible that the distribution of performance may have been impacted in a way that would balance out the average effect (e.g., improved performance at the bottom of the distribution and worse performance at the bottom of the distribution). However, the distribution of test scores across groups shows no significant heterogeneity by level of performance except for a small range around the average performance (Figure 6).

Teaching practices improved slightly in the WSD group. As Table 10 shows, the probability that the teacher frequently used the blackboard increased by 7% relative to the control group and teachers were more likely (10%) to call on student by their names (both results significant with 90% confidence). However, we see no evidence that the program affected the confidence of children to participate and ask questions during class. Similarly, the programs did not improve the likelihood that a teacher would prepare for the class with written notes.

The first four columns in Table 11 indicate that the intervention groups are more likely than the control group to consult teachers, parents, and the regional office for planning and decisions about school expenses. The point estimates in column 4 indicate that the WSD group relies less on the regional education authorities than the Grant-only group, potentially due to the training component of the WSD. Moreover, the WSD group is more

likely to conduct fundraisers relative to the control group, whereas this is not the case for the Grant-only group. The WSD treatment has a negative effect on the number of overall PTA meetings: On average, PTAs met 0.41 less in the WSD group than in the Control group (column 7, Table 11). The likely explanation for this finding may be the fact that the WSD creates six sub-committees (as observed in the one-year follow-up data) within the community to deal with different challenges pertaining to the functioning of the school. Parents may participate in sub-committee meetings and so the school may hold fewer overall PTA meetings. Although some of the changes observed may be expected to impact student learning, we observe no impact on student performance (Table 12).

6. Final results

6.1. Average Treatment Effects on learning outcomes and participation

The main outcome variables of interest are the learning outcomes measured by a comprehensive written test. Other outcomes of interest beside student test scores include measures of absenteeism for teachers and for students, and a measure of enrollment. Table 13 presents the estimates of Equation 1 where the dependent variable is a standardized test score in math or English. The estimates show that the interventions have no positive effect on student math and English test scores. The point estimates are mostly negative but small and statistically insignificant. A test comparing the mean score between the WSD and the Grant-only does not reject the null hypothesis that the two interventions have the same effect on test scores.[15]

---

[15] Across both treatment groups, school identified the largest budget item on which the grant was spent: 46% reported teaching and learning materials (including stationery), 23% reported infrastructure (e.g., furniture, building improvements), 20% reported some kind of workshop, 7% reported a radio, while a few reported spending the grant on garden materials.

Table **14** presents the same results, controlling for baseline test scores: The significant patterns are identical and all coefficients of interest are within 0.02 of each other.

We run the same model where the outcome variables are student absenteeism and teacher absenteeism. The estimates in the first column of Table 15 indicate that the WSD intervention reduced student absenteeism by about 5 percentage points from a base of about 23% (significant at the 5% level). This corresponds to a nearly 21% reduction in absenteeism. The second column shows that the WSD reduced teacher absenteeism by about 3 percentage points from a base of about 13%, which represents a 23% reduction in teacher absenteeism. We observe no impact of student enrollment.

6.2. Discussion, Interpretation, and potential mechanisms

The Whole School Development program, over time, had a positive impact on student and teacher school attendance. In theory, increased participation should translate into increased learning outcomes. However, in this case we observe increased participation but no change in test scores. We explore four potential explanations for this finding: (1) Selection, (2) Poor teacher quality, (3) Human capital in the community, and (4) Improvements in the control schools.

6.2.1. Selection as treatment effect

One plausible explanation could be that the increased student participation brought back students that perform poorer than the average. If the intervention has brought in worse-performing students in the intervention group, then the average treatment effect (ATE) may be biased downward. The distribution of test scores shown in Figure 7 shows a shift to the left, albeit only at the left tail. This is suggestive evidence for the hypothesis that the WSD program attracted more low-performing students into schools. Miguel and Kremer's (2004) analysis of a de-worming intervention in Kenya found large effects on participation but no effect on test scores: this same kind of selection was a potential explanation in that context as well.

If students who attended more because of the WSD were also students who otherwise perform more poorly, then one might expect the treatment effect to be larger at higher percentiles of the performance distribution. To verify this, we first look at the treatment effect in the each quantile. Figure 8 shows an upward trend, which partially supports this story. However, for this effect to be interpreted as the effect of the intervention on the students on the respective quintiles, the rank preserving assumption between the baseline and the end-line needs to be true. (In other words, one must assume that students would occupy the same rank in the test score distribution independent of the intervention: If Student A had better scores than Student B before the intervention, she would continue to outperform Student B after the intervention, even though one or both of them may have improved.) This is a strong assumption and there is no way to test it.

Nevertheless, we address this selection issue by bounding the treatment effect using Lee's trimming procedure (Lee 2009). The procedure consists of dropping a proportion of the lower tail of the distribution in the WSD group – i.e., those low-performing students drawn to schools by the intervention – in order to construct an upper bound of the effect of the intervention. Then we drop a proportion of the upper tail in order to construct a lower bound. Lee shows that the proportion to trim is given by

$$p = \frac{\% \, present_{WSD} - \% \, present_{CONTROL}}{\% \, present_{WSD}}$$

Let y_x be the test score of student i and $y_x = G^{-1}(x)$ with G being the cumulative distribution function of y conditional on being in the WSD group and being successfully tracked. Then, the sharpest bounds of the treatment effect are given by calculating the sample counterpart of the following:

$$\mu_{Upper \, Bound} = E[y|WSD, Present, y \geq y_p] - E[y|Control, Present]$$

and

$$\mu_{Lower \, Bound} = E[y|WSD, Present, y \leq y_{1-p}] - E[y|Control, Present]$$

Under the assumption of independence and monotonicity, these bounds are shown to be the smallest upper bound and the largest lower bounds that are consistent with the data at hand. The bounds can be calculated only using the subset of students that we tracked by design from the baseline to the end line. These students were five third-graders per school in 2008 who were in sixth grade at the end. At the end, we were able to find 71% of them in the control schools versus 79% in the WSD schools. The average test scores are comparable between the two groups, but if the extra students tracked in the WSD are weaker on average, then this comparison will be biased in favor of not finding an effect (

Table **16**).

Table **17** presents the estimates of Lee's sharp bounds, accounting for selection. The results indicate an upper bound of 0.17 and a lower bound of -0.19 standard deviations on the mathematics test score. The effect on English is bounded by 0.26 and -0.16 standard deviations. These ranges are not a confidence interval for the average treatment effect, but a range of point estimates that are all consistent with the data given the selection concern. Given these bounds (which clearly include a zero effect), and given the underlying assumption on the absentees, it is reasonable to lean toward an interpretation

of no significant effect. These findings suggest that the selection issue may not be pronounced.[16]

6.2.2. Poor complementary inputs: Teacher quality

A third explanation for a lack of learning effects in the face of attendance improvements is that other inputs such as teacher quality are sufficiently low that increased participation does not translate into improved learning outcomes. In 2009, we conducted a teacher content knowledge test. The test consisted of the same test applied to students, with a few additional questions from Gambian secondary school textbooks. Figure 9: Teacher Content knowledge on selected English questions and Figure 10 show sample questions from the test and average teacher performance on them. The findings suggest that teacher content knowledge was indeed low: Only 2.6% of teachers scored 95% or more, and over one-third of the teachers scored below 75%. There were no significant differences across treatment groups.

Figure 11 shows a positive correlation between matched teacher and pupil test scores.[17] Sixth-grade math test scores mainly drive the correlation. In addition, the result from classroom observation indicates that only about 45% of the instructional time is actually focused on learning activities (Table 18), to be contrasted with estimates between 52% and 65% in a sample of Latin American countries (Bruns & Luque 2014). Taken together, these results suggest that teacher quality and effectiveness may be so low in the Gambia that other school improvement interventions will be ineffective.

6.2.3. Community human capital at baseline: Heterogeneity

The Gambia is characterized by a low adult literacy rate, especially in rural areas. This characteristic was reflected in the School Management Committees. Nearly 4 out of 5 committee members from the community (i.e., not school employees) had no formal education and only 16% had completed at least primary education. Some level of human capital may be needed at the local level for interventions such as the WSD to build on. For example, for parents to effectively help to run the school, the parents would need some schooling of their own. We investigated this hypothesis by interacting the interventions with a baseline measure of human capital.

$$Outcome_{isd} = \alpha + \beta_1 WSD_{sd} + \beta_2 Grant_{sd} + \beta_3 Baseline\ Human\ Capital_d +$$

$$\beta_4 Grant_{sd} * Baseline\ Human\ Capital_d + \beta_5 WSD_{sd} * Baseline\ Human\ Capital_d + \varepsilon_{isd} \quad (2)$$

---

[16] Note that these bounds do not account for the potential peer effect from absentees that are coming back, i.e., if poorer performing students were returning and not only bringing down the average test scores but negatively affecting the performance of student who were previously attending. To account for this particular aspect, one would need a structural model, which is beyond the scope of this paper.

[17] This could of course in part be driven by selection, if higher-performing students are placed in the classes of higher-performing teachers.

We report estimates of equation 2 in

Table 19, where *BaselineHC_d* is the district level adult literacy in 2006. Across the districts included in the evaluation, the average adult literacy was 31%, ranging from 12% to 53% across the localities where the schools are located. The interaction between WSD and adult literacy in 2006 has a significant and positive effect on both math and English test scores. This suggests that human capital, at least measured as adult literacy, has an amplifying effect on the WSD. The same is not found for the Grant-only intervention.

The estimates also suggest that interventions such as WSD could potentially have detrimental effects in places where human capital is very low. One channel for this negative effect could be that shifting from one set of management practices to another is costly. If existing practices are functioning at some level and new practices (which are expected to be better) are not properly adopted, the end outcome could be negative. Furthermore, WSD shifts some degree of decision making from school leaders to the community: If the community has very little capacity, then the result on school management quality could be negative. This is also consistent with the multitasking literature (e.g., Holmstrom and Milgrom 1991), which – in this case – suggests that when asked to perform many tasks simultaneously (as in an integrated program such as WSD), schools would prioritize some tasks over others. However, if the different tasks are complements, then improvements in just a few may not yield a positive overall outcome. Table 20 presents the same estimates where *BaselineHC_d* is replaced by the percentage of the school management committee members who have no formal education. The results are qualitatively the same.

We graphically present the results of this analysis in Figures Figure 12 and Figure 13. We conclude that the WSD intervention is likely to improve learning outcomes in areas with high baseline human capital, but it could be counterproductive in areas where the basic human capital is very low. Our point estimates suggest that the WSD would have a positive impact on learning outcomes if the level of adult literacy at the baseline were greater than 45%.

To further understand this human capital aspect, we also conducted qualitative analysis. After two years of exposure to the WSD program, we asked the head teachers about their opinion regarding shifting school management to the schools and the communities. Most of the head teachers (75%) disapproved of this idea, 19% thought that it would be a good idea and, 6% expressed no opinion either way. Most of the head teachers who approved the idea supported their position with the argument that the communities and the schools better know their problems and that it would be more effective to allow them to handle them. Others pointed out that it would induce more accountability as the teacher can be monitored more effectively and action can be taken in a timely fashion if they do not deliver.

However, most head teachers disagree with that point of view in the context of Gambia. Almost all of those who opposed the idea pointed out the lack of capacity at the local level to manage the school. As one head teacher expressed, such decentralized decision

making would be "almost impossible because a large portion of our communities are illiterate."

Even though standards are low, pupils are performing poorly, and teacher content knowledge is problematic, over 90% of parents are satisfied with the school and think that the school is doing fine in training their children. When asked to give the reason why they make such assessments, 83% of the parents say that the child is performing well and that the school has good teachers. Another 15% based their assessment on the fact that the child is better behaved and disciplined at home.

Similarly, over 90% of the parents report high aspirations for their children. They reported wanting them to study to the highest level and enter careers with high social esteem such as doctors, ministers, etc. These responses indicate that these parents care about and value the educational outcomes of their children, but there is a contrast between this aspiration and their ability to assess the effectiveness of the school and hold the teachers accountable.

This large disconnect between actual student academic performance (and, consequently, school performance) and the parents' assessment is in tune with the theoretical motivation of this paper. Among the few parents who are dissatisfied by student and school performance, most pointed out specifics about the incapacity of the child to read and write properly, and the mismanagement of the school. It may be that those parents are more educated and better able to assess the progress of the children and the performance of the school. These findings confirm that the WSD intervention may be more appropriate where local capacity is sufficiently high.

In addition to increasing the capacity of communities to hold schools accountable, district-level literacy could be a proxy for some other effect. One possibility is that the literacy effect, demonstrated in Table 19 and Table 20, is actually a proxy for some other, correlated characteristic of households or districts. We test this by interacting the WSD treatment with socio-economic status at the household and the district level in Tables Table **21** and

Table 22; those interactions are not statistically significant. Another possibility is that adult literacy is a proxy for some baseline level of ability among the students. We test this by interacting WSD with baseline test scores (Table

Table 23) and again find no significant relationship. A third possibility is that, rather than having additional management capacity, more literate districts are wealthier and are contributing more to schools. In

Table 24 we present evidence that higher economic status does not associate with higher likelihood of financial or in-kind contributions to the school. Taken together, these tests further suggest that the human capital effect is in effect a proxy for the capacity of parents within the district.

6.2.4. Improvement in the control schools?

The lack of impact on average test scores could also be due to the fact that control schools have improved as well, through mechanisms other than increased participation. Since the school management manuals were made available to all the schools, it is possible that the control group would implement at least part of the practices, although it seems unlikely that they would have adopted a similar set of practices to the WSD schools without any support. Qualitatively, we found no evidence that they used the manual. In addition, our test score data from 2008 and 2010 were collected at the same grade level. This allows us to conduct a before and after analysis in each group, including the control group (25). We find no evidence of a positive time trend in the control group between the baseline and the 2010 test scores.

8. Conclusion and future research

In this research, we evaluated a school management training program in the Gambia called Whole School Development (WSD). Intermediate results one year post-intervention showed some basic changes in school organization in the WSD schools but no effect on test scores or on student and teacher absenteeism. These results served mostly as evidence of project implementation. Two years post-intervention, we found no effect on test scores but modest positive effects on student and teacher participation measured by the prevalence of absenteeism.

Three years into the program, we found no effect of the WSD intervention on learning outcomes measured by scores on a comprehensive test of Mathematics and English. However, we found a large effect on participation: The intervention led to reductions in student and teacher absenteeism respectively of nearly 5 percentage points from a base of 24%, and about 3 percentage points from a base of about 13%. We found no effect of the Grant-only intervention relative to the control on test scores or on participation.

Since this intervention emphasized local capacity building, we analyzed the heterogeneity of the effectiveness of the program by one dimension of initial capacity, adult literacy. Our findings suggest that the WSD may be effective when adult literacy at baseline is sufficiently high. The range of the estimated effects suggests that, for places where local capacity is extremely low, this intervention could potentially be counterproductive as the reform may shift decision making away from school leaders with relatively higher human capital.

We also observed a large disconnect between the parents' evaluation and the actual performance of the schools. Whereas evidence from student tests reveals poor

performance of children, over 90% of the parents are satisfied with the schools and their children's performance. This disconnect may explain why parents do not hold the schools accountable and participate effectively in school management. Parents have very high professional aspirations for their children, but the evidence suggests that they may lack the ability to evaluate the performance of their children and thus to demand accountability from educators. That is precisely what the capacity building component of the WSD attempted to address, but the WSD does not appear to have accomplished this, at least not sufficiently to change test scores. While the WSD focused on concrete actions by parents to hold schools accountable, the relevant challenge may be more related to the basic inability of parents to read and write.

With the grant-only intervention, we found no evidence of positive effects on outcomes, except on process variables such as community engagement in decision making. However, there are many reasons why this should be taken with caution. First, principals found the disbursement process cumbersome because disbursements had to be approved by the regional directorates. This may have prevented schools from effectively addressing issues that required immediate attention. Second, and perhaps most importantly, the one-time grant was relatively small to expect a substantial effect three years later (although note that no effect was observed at any point). With an increased amount or with more sustained yearly grants, the results might differ.

Based on this study, we draw the following conclusions and policy implications. First, a crucial feature for an effective local management program, such as the one envisioned and studied here, is local human capital (such as literacy) in the communities. We hypothesize that in general, the gap between capacity at the central and local levels is a key determinant of the success of such policies. In countries where this gap is small, regardless of absolute capacity levels, a decentralized policy may be superior because of the added value of localized information. However, if the gap is sufficiently high in favor of the central government, then localized information is less useful because communities are not well equipped to act on it. Our findings show that the Gambia may fall in the latter group. An intervention like this one may not be effectively by itself for the median community. Rather, interventions to increase community involvement should seek to relax constraints on community capacity.

Second, in the Gambia, there appear to be other binding constraints on the education production function. Two of these constraints, explored here, are teacher capacity and effectiveness; others are limited instructional time due to the widespread double-shift schools, and teacher compensation. National policy shifts may need to lay the groundwork for improvements in these areas before school-level improvement plans can be effective.

Third, our findings suggest that a mechanism to supply accurate information to communities (about the relative performance of their children and the schools) could be desirable. This, in essence, substitutes for baseline capacity on the part of parents to

evaluate the schools. Our data suggest that most parents – including in the rural areas – have high aspirations for their children's professional futures and educational achievements. However, this is juxtaposed with the sharp inability of parents to understand the performance of their children and the functioning of the schools, even after the intervention. If well informed, parents may seek to hold schools accountable for their children's learning outcomes. In recent years, the government has experimented with providing school report cards that are focused on pictograms (such as smiley faces). Such a communication intervention that does not require high levels of literacy would be worth testing.

Our findings call for nuance in the design of policies that decentralize school management to communities. School-based management is gaining popularity in low-income countries (Barrera-Osorio et al. 2009; Bruns et. al. 2011). In Africa alone, there are many ongoing field experiments to test variants of school-based management policies. These studies will shed much needed light on which models will help communities to keep schools accountable.

Works Cited

Abadzi, Helen. Efficient learning for the poor: New insights into literacy acquisition for children. *International Review of Education*, 54:5-6, November 2008.

Adekanmbi, Adebimpe, Moussa P. Blimpo, and David K. Evans. The state of lower basic education in the Gambia: A baseline survey report prepared for the ministry for basic & secondary education, the Gambia. *APEIE, World Bank*. Available upon request to authors, 2009.

Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. Pitfalls of participatory programs: evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, 2:1–30, 2010.

Bardhan, Pranab, and Dilip Mookherjee. Relative capture of local and central government: An essay in the political economy of decentralization. Center for International and Development Economics Research, Institute of Business and Economic Research, UC Berkeley, 2002.

Barrera-Osorio, Felipe, Tazeen Fasih, and Harry Anthony Patrinos with Lucrecia Santibez. Decentralized Decision-Making in Schools: The Theory and Evidence on School- Based Management. 2009.

Bayona, E.L.M and B. Sadiki. An investigation into appropriate ways of implementing institutional development (whole school development). University of Venda School of Education research report, *Accessed at http://www.jet.org.za/publications/pei-research* on 31 July 2010, 1999.

Bjorkman, Martina, and Jakob Svensson. Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124:735–769, 2009.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. Does Management Matter in Schools? Working paper. 2014.

Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos. Making Schools Work: New Evidence on Accountability Reforms. World Bank. 2011.

Bruns, Barbara, and Javier Luque. Great Teachers: How to Raise Student Learning in Latin America and the Caribbean. World Bank. 2014.

Conn, Katharine, Esther Duflo, Pascaline Dupas, Michael Kremer, and Owen Ozier. Bursary targeting strategies: Which method(s) most effectively identify the poorest primary

school students for secondary school bursaries? Innovations for Poverty Action Kenya, *Unpublished Report*, 2008.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. 2014.

Gertler, Paul, Harry Patrinos, and Marta Rubio-Bodina. Empowering parents to improve education: Evidence from rural Mexico. Journal of Development Economics 99(1), 2012.

Gugerty, Mary Kay and Michael Kremer. Outside funding and dynamics of participation in community associations. *American Journal of Political Science*, 52:585–602, 2008.

Hanushek, Eric A. and Ludger Woessmann. The role of education quality for economic growth. *World Bank Policy Research Working Paper* No. 4122, 2007.

Hanushek, Eric A. and Ludger Woessmann. Schooling, cognitive skills, and the Latin American growth puzzle. *NBER Working Paper* 15066, 2009.

Holmstrom, Bengt, and Paul Milgrom. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7:24-52, 1991.

Jimenez, Emmanuel and Yasuyuki Sawada. Do community-managed schools work? an evaluation of El Salvador's educo program. *The World Bank Economic Review*, 13:415–441, 1999.

Lee, David S. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, pages 1071 – 1102, 2009.

Miguel, Edward and Michael Kremer. Worms: Identifying impacts on educational and health in the presence of treatment externality. *Econometrica*, 72:159 – 217, 2004.

Pugatch, Todd and Elizabeth Schroeder. Incentives for Teacher Relocation: Evidence from the Gambian Hardship Allowance. *Economics of Education Review,* 41:120-136, 2014.

Reinikka, Ritva and Jakob Svensson. Local capture: Evidences from a central govern- ment transfer program in uganda. *The Quarterly Journal of Economics*, 2004.

USAID, The World Bank, & Eddata. *The Gambia Early Grade Reading Assessment (EGRA): Results from the 1,200 Gambian Primary Students Learning to Read in English—Report for the World Bank.* January 2008.

World Bank. Making Services Work for Poor People. *World Development Report*, 2004.

Tables

Table 1: Key elements of intervention arms

	Grant provided	Management training provided	Management manual provided
WSD Schools	Yes	Yes	Yes
Grant-only Schools	Yes	No	Yes
Control Schools	No	No	Yes

Table 2: Timeline of intervention and evaluation

Date	Activity
10/2007 – 4/2008	Sensitization and coordination between stakeholders
4/2008 – 6/2008	Assignment to interventions and baseline data collection
5/2008 – 12/2008	Grant distribution and training in the WSD schools
5/2009 – 6/2009	Collection of first follow-up data
5/2010 – 6/2010	Collection of second follow-up data
5/2011 – 6/2011	Collection of third follow-up data
Throughout	Monitoring of implementation

Table 3: Description of the data

Year	Data type	Respondent	Obs	Notes
2008	School data	Principal, deputy	273	
	Student test	3rd & 5th grades	8,856	
	Classroom visit	4th & 6th grades	528	
	Student interview	3rd & 5th grades	2,688	Administered to a subset of tested children
2009	School data	Principal, deputy	176	
	Student test	4th & 6th grades	5,660	
	Classroom visit	3rd & 5th grades	346	No data in Grant-only schools
	Student interview	4th & 6th grades	1,755	
	Teacher test	About 6 teachers	1,049	
2010	School data	Principal, deputy	276	
	Student test	3rd & 5th grades	9,022	
	Classroom visit	4th & 6th grades	502	
	Student interview	3rd & 5th grades	2,678	
	Parent interview	Parent or caregiver	567	Of two interviewed students
2011	School data	Principal, deputy	274	
	Student test	4th & 6th grades	5,230	
	Classroom visit	3rd & 5th grades	534	
	Student interview	4th & 6th grades	2,579	
	SMC interview	Committee (minus principal)	249	Mostly PTAs, in controls and Grant
	Teacher interview	4th & 6th grades	517	Teachers of tested students

Table 4: Baseline Group Comparison on School Characteristics

	WSD	Grant	Control
Student Observations			
Number of students	461	433	426
	(59)	(41)	(45)
Student-teacher ratio	32	34	32
	(0.89)	(0.97)	(1.14)
Double shift	0.33	0.49	0.41
	(0.50)	(0.50)	(0.05)
Tap drinking water	0.23	0.20*	0.33
	(0.04)	(0.04)	(0.05)
Student-latrine ratio	79	49	64
	(15)	(4)	(9)
Has a library/storage for books	0.37	0.53	0.47
	(0.05)	(0.05)	(0.05)
Received cash/in-kind from community	0.38	0.31	0.29
	(0.05)	(0.05)	(0.05)
Number of meetings with parents	4.39**	3.70	3.69
	(0.27)	(0.24)	(0.25)
Has mentoring system	0.86	0.82	0.81
	(0.04)	(0.04)	(0.04)
Written staff code of conduct	0.39	0.43	0.44
	(0.05)	(0.05)	(0.05)
Pupils per class (2006 Administrative Data)	34	33	34
	(0.10)	(0.10)	(0.11)
Adult literacy (2003 Census)	38%	39%	38%
	(0.015)	(0.014)	(0.012)
Primary Education or more (2003 Census)	57%	55%	55%
	(0.017)	(0.016)	(0.014)
Years Established	24	25	24
	(1.6)	(1.8)	(1.9)
Number of observations	90	94	89
Classroom Observations			
Teacher has lesson notes	0.31	0.33	0.27
	(0.04)	(0.04)	(0.03)
Percentage of pupils absent	0.25	0.21*	0.26
	(0.06)	(0.02)	(0.02)
Hours/week English	3.67	3.57	3.81
	(0.15)	(0.15)	(0.13)
Number of observations	175	180	173

Notes: Standard errors are in parentheses. *** 1% Significance Level, **5% significance Level, *10% Significance Level. The mean comparison test contrasts each treatment group with the control group.

Table 5: Baseline Group Comparison on Student Characteristics

	3rd grade			5th grade		
	WSD	Grant	Control	WSD	Grant	Control
Student age	10.20	10.20	10.10	12.73	12.59	12.64
	(0.10)	(0.10)	(0.10)	(0.08)	(0.08)	(0.08)
Number of siblings	4.90	4.70	4.75	4.70	4.70	4.80
	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)	(0.12)
Ate breakfast today	0.69	0.71	0.73	0.67**	0.73	0.74
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ate lunch yesterday	0.96	0.95	0.94	0.94	0.97	0.95
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Electricity at home	0.19*	0.21	0.24	0.20	0.17	0.20
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Radio at home	0.91	0.92	0.93	0.88	0.89	0.87
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
TV at home	0.37	0.38	0.38	0.40	0.36	0.36
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Telephone/Mobile at home	0.83	0.81	0.82	0.81	0.86	0.83
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Percent repeating the Class	0.09	0.09	0.09	0.08	0.07	0.08
	(0.29)	(0.29)	(0.29)	(0.26)	(0.26)	(0.26)
Observations	462	462	445	423	458	447

Notes: Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The mean comparison test contrasts each treatment group with the control group.

Table 6: Community participation, school management and characteristics (2009)

	WSD	Control	Difference	P-value
Received support/aid from the community	0.46	0.35	0.11	0.15
	(0.05)	(0.05)	(0.07)	
Does the school have a PTA	1.0	0.99	0.01	0.32
	(0)	(0.01)	(0.01)	
PTA fund raisers	0.10	0.11	-0.01	0.83
	(0.03)	(0.03)	(0.05)	
PTA member contribution	0.09	0.05	0.04	0.23
	(0.03)	(0.02)	(0.04)	
PTA not funded	0.71	0.75	-0.04	0.57
	(0.05)	(0.05)	(0.07	
Number of meetings with the parents or PTA	4.45	3.92	0.53	0.19
	(0.31)	(0.26)	(0.4)	
Mentoring system in place for junior teachers	0.47	0.53	-0.06	0.41
	(0.05)	(0.05)	(0.08)	
Mentors trained	0.7	0.57	0.14*	0.08
	(0.05)	(0.05)	(0.08)	
Leadership and Management committee in place	0.94	0.75	0.19***	0
	(0.03)	(0.06)	(0.06)	
Community Participation committee in place	0.79	0.63	0.16**	0.04
	(0.05)	(0.07)	(0.08)	
Curriculum Management committee in place	0.84	0.51	0.33***	0
	(0.04)	(0.07)	(0.08)	
Teachers' professional development com. in place	0.8	0.61	0.19**	0.02
	(0.05)	(0.07)	(0.08)	
Teaching and learning resources com. in place	0.81	0.59	0.22**	0.01
	(0.05)	(0.07)	(0.08)	
Learners welfare committee in place	0.88	0.71	0.17**	0.01
	(0.04)	(0.06)	(0.07)	
School has developed school policy	0.45	0.36	0.09	0.26
	(0.05)	(0.05)	(0.07)	
First grade enrollment	91.82	76.29	15.53	0.2
	(9.85)	(7.02)	(12.12)	
Student-teacher ratio (Lower Basic)	53.18	53.18	0	1
	(11.55)	(7)	(13.11)	
Seen records of the teachers attendance	0.91	0.89	0.02	0.64
	(0.03)	(0.03)	(0.05)	
Teacher Absenteeism/ Average 5 random days	0.06	0.06	0	0.94
	(0.01)	(0.01)	(0.01)	
School has a library	0.53	0.6	-0.07	0.43
	(0.05)	(0.05)	(0.08)	
Observations	88	89		

Notes: Standard deviations in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance level. The test of comparison of means is between each treatment group and the control group.

Table 7: Teaching practices and absenteeism (First follow-up in 2009)

	WSD	Control	Difference	P-value
Teacher absent (at our arrival)	0.11	0.12	0.01	0.73
	(0.02)	(0.03)	(0.04)	
Teacher missed at least one day last week	0.26	0.33	0.07	0.16
	(0.03)	(0.04)	(0.05)	
Teacher Absenteeism (Five random days average)	0.06	0.06	0	0.94
	(0.01)	(0.01)	(0.01)	
Student Absenteeism (Day of test)	0.26	0.24	0.02	0.55
	(0.02	(0.01	(0.02	
Student Absenteeism (Five random days average)	0.38	0.36	0.02	0.71
	(0.04)	(0.03)	(0.05)	
Teacher has written lesson plan	0.56	0.67	-0.11**	0.04
	(0.04)	(0.04)	(0.05)	
Teacher has a written lesson note for today's lesson	0.32	0.41	-0.09*	0.08
	(0.04)	(0.04)	(0.05)	
Teacher missed at least one day last week	0.26	0.33	0.07	0.16
	(0.03)	(0.04)	(0.05)	
Call out children by their names	0.48	0.35	0.13**	0.03
	(0.04)	(0.04)	(0.06)	
Address questions to the children during class	0.69	0.75	0.06	0.27
	(0.04)	(0.04)	(0.05)	
Encourages the children to participate	0.61	0.68	0.07	0.23
	(0.04)	(0.04)	(0.06)	
The children used textbooks during the class	0.38	0.47	-0.09*	0.09
	(0.04)	(0.04)	(0.05)	
The children used workbooks during the class	0.54	0.45	0.08	0.14
	(0.04)	(0.04)	(0.06)	
The children ask questions for clarification	0.26	0.23	0.03	
	(0.04)	(0.03)	(0.05)	
Observations	88/169	89/177		

Notes: Standard deviations in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. Based on data from school and classroom visits.

Table 8: Student performance (First follow-up in 2009)

	4th Grade			6th Grade		
	WSD	Control	P-value	WSD	Control	P-value
Reading test						
Correct letters per minute	55	57	0.26	73	75	0.17
	(1.23)	(1.23)		(1.15)	(1.1)	
Correct words per minute	23	25	0.33	41	41	0.75
	(1.18)	(1.15)		(1.08)	(1)	
Written test						
Overall	47.2	48.22	0.5	60.59	61.79	0.4
	(0.46)	(0.45)		(0.49)	(0.45)	
Math	47.04	49.75	0.2	65.95	68.19	0.23
	(0.65)	(0.66)		(0.67)	(0.62)	
Literacy	45.82	45.94	0.93	57.19	57.76	0.67
	(0.44)	(0.41)		(0.47)	(0.43)	
Observations	411	403		431	460	

Notes: Standard deviations in parentheses. *** 1% Significance Level, **5% Significance Level,*10% Significance Level. Same students at the baseline. The score of the written test is the average score expressed in percentage.

Table 9: Student performance (two years into intervention – 2010)

	Test score [a]		Percentage of students who can read [b]	
	3rd graders	5th graders	3rd graders	5th graders
WSD	-0.001	-0.08	0.01	-0.05
	(0.08)	(0.09)	(0.03)	(0.04)
Grant	0.01	0.03	-0.01	-0.05
	(0.08)	(0.09)	(0.02)	(0.04)
Observations	4537	4354	1241	1202
Mean of dependent variable in comparison group	35.32% [a]	52.06% [a]	11% [b]	38% [b]

Notes: Standard deviations in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. [a] Test score is normalized to 100 points. It is standardized only for the calculation of the treatment effect. [b] Percentage of students who can read 45 or more words per minute.

Table 10: Teaching practices (two years into intervention – 2010)

	Probability of calling students by name	Probability of frequent use of the blackboard	Probability of children asking questions in class	Probability that the teacher has *no* lesson notes
	(1)	(2)	(3)	(4)
WSD	0.10*	0.07*	0.03	0.03
	(0.07)	(0.03)	(0.06)	(0.06)
Grant	-0.001	0.02	-0.08	-0.01
	(0.07)	(0.04)	(0.06)	(0.06)
Observations	427	427	420	511
Mean of dependent variable in comparison group [a]	39%	82%	33%	37%

Notes: Standard deviations in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The unit of observation is a classroom. Robust standard errors. All coefficients are marginal probabilities. [a] Percent of classrooms where dependent variable is 1.

Table 11: Participation in management (two years into intervention –2010)

	Marginal Probability to participate in decision-making						
	Teachers	Parent	Rely on SDP	RED	Fundraisers	Know PTA memb.	# Meetings parent/school
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	Probit	Probit	Probit	Probit	Probit	Probit	OLS
WSD	0.42***	0.64***	0.18***	0.26***	0.11**	-0.15**	-0.41***
	(0.08)	(0.06)	(0.07)	(0.08)	(0.06)	(0.08)	(0.18)
Grant	0.37***	0.65***	0.16**	0.37***	0.07	-0.04	-0.26
	(0.08)	(0.06)	(0.07)	(0.08)	(0.06)	(0.08)	(0.18)
Observations	274	274	274	274	274	505	505
Mean of dependent variable in comparison	3.3%	9%	1%	2%	7%	50%	1.9

Notes: Marginal effects are reported for Probit regressions. Robust standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The unit of observation is the school in the first four columns and the household in the remaining columns. RED = Regional Education Directorate. SDP = School Development Plan.

Table 12: Treatment effect on student performance and learning outcomes – two years into intervention (2010)

	3rd Grade		5th Grade	
	Standardized test score [a]	Probability that a child can read 45 or more words per minute	Standardized test score [a]	Percentage of students who can read 45 or more words per minute
WSD group	-0.001	0.01	-0.08	-0.05
	(0.08)	(0.03)	(0.09)	(0.04)
GRANT group	0.01	-0.01	0.03	-0.05
	(0.08)	(0.02)	(0.09)	(0.04)
Number of observations	4537	1241	4354	1202
Mean of dependent variable in comparison group	35.32%	11%	52.06%	38%

Notes: Robust standard error clustered at school level in parenthesis. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. [a] Test score normalized to 100 point. It is standardized only for the calculation of the treatment effect.

Table 13: Average Treatment Effect on 4th and 6th-Graders – Three to four years into the intervention

	Math	English
WSD	-0.05	0.01
	(0.07)	(0.08)
Grant	-0.07	-0.08
	(0.06)	(0.07)
4th Grade Dummy	-0.69***	-0.74***
	(0.03)	(0.03)
Constant	0.40***	0.42***
	(0.04)	(0.05)
P-value WSD = Grant	0.76	0.23
Observations	4817	4817

Notes: Standard errors in parentheses. *** 1% Significance Level, **5%. Significance Level, *10% Significance Level.

Table 14: Average Treatment Effect on 4ᵗʰ and 6th-Graders – Three to four years into the intervention – Controlling for variables at baseline

	Math	English
WSD	-0.06	0.00
	(0.07)	(0.06)
Grant	-0.07	-0.06
	(0.06)	(0.06)
Baseline PTA Meetings	-0.01	-0.01
	(0.01)	(0.01)
Baseline Test Scores	0.34 ***	0.46***
	(0.05)	(0.06)
4th Grade Dummy	-0.68***	-0.73***
	(0.03)	(0.03)
Constant	-0.21***	-0.27***
	(0.07)	(0.07)
P-value WSD = Grant	0.92	0.30
Observations	4716	4716

Notes: Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level.

Table 15: Effect of the Interventions on Student and Teacher Absenteeism, and on Enrollment

	Absenteeism		Log First-Grade
	Students	Teachers	Enrollment
WSD	-4.94**	-3.11*	-0.01
	(2.24)	(1.75)	(0.1)
Grant	-2.61	-0.22	0.03
	(2.24)	(1.76)	(0.1)
Constant	23.35***	13.31***	4.16***
	(1.72)	(0.01)	(1.26)
P-value WSD = Grant	0.25	0.11	0.62
Observations	407	274	274

Notes: Robust standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The dependent variable in the first column is the percentage of student absent on the day of survey (scale of 0-100). The dependent variable in the second column is percentage of teachers absent (scale of 0 - 100). The dependent variable in the third column is the log enrollment of first-graders. The unit of observation in the first column is the classroom. The unit

Table 16: Inputs to Lee trimming procedure

	Control	Treatment
Number of observations	444	453
Proportion non-missing	71.0%	79.3%
Math score	73.0%	71.1%
	(20)	(23)
English score	61.0%	62.0%
	(18)	(21)

Notes: The dependent variable is a standardized test score. Standard deviations are in parentheses.

Table 17: Bounds for the average treatment effect, accounting for selection using the trimming procedure

	Lee's upper bound	Lee's lower bound
Math	0.17	-0.19
	(0.06)	(0.09)
English	0.26	-0.16
	(0.07)	(0.11)

Notes: The dependent variable is a standardized test score. Standard errors are in parentheses.

Table 18: Classroom Stallings, instructional time allocation

	Share of time* (%)			
	All	WSD	GRANT	CONTROL
Learning activities	44	44	44	45
Social interaction	22	21	23	22
Student (s) uninvolved	19	20	18	19
Discipline	1	1	2	1
Classroom management	2	2	1	1
Classroom management alone	3	3	3	2
Teacher out of the room	9	8	10	10
Obs.	534	176	183	175

Notes: Based on ten two-minute snapshots of classroom activities in 534 classroom observations.

Table 19: Role of baseline levels of human capital

	Math	English
WSD	-0.50***	-0.31*
	(0.17)	(0.17)
Grant	-0.13	0.01
	(0.16)	(0.18)
Adult Literacy	0.54*	1.66***
	(0.32)	(0.37)
WSD × Adult Literacy	1.12**	0.78*
	(0.46)	(0.51)
Grant × Adult Literacy	0.07	-0.46
	(0.43)	(0.54)
Constant	0.25	-0.10
	(0.11)	(0.12)
Observations	2331	2331

Notes: Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level,*10% Significance Level. Adult literacy is the district level percentage of adults who are literate. It is expressed in the range 0-1.

Table 20: Role of human capital at the baseline

	Math	English
WSD	0.36	0.38
	(0.24)	(0.28)
Grant	0.17	0.20
	(0.25)	(0.32)
SMC Literacy	0.02	-0.28
	(0.21)	(0.24)
WSD × SMC Illiteracy	-0.65**	-0.57*
	(0.29)	(0.34)
Grant × SMC Illiteracy	-0.36	-0.39
	(0.30)	(0.39)
Constant	0.41	0.64
	(0.17)	(0.21)
Observations	2035	2035

Notes: Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. SMC illiteracy is the percentage of the School Management Committee members who have no formal education. It is expressed in the range 0-1.

Table 21: The effect of baseline socio-economic status (using individual SES)

	Math	English
WSD	-0.14*	-0.08
	(0.07)	(0.08)
Grant	-0.14*	-0.17*
	(0.08)	(0.08)
Child's SES	0.07*	0.05
	(0.04)	(0.04)
WSD × Child's 2011 SES	-0.07	0.04
	(0.06)	(0.34)
Grant × Child's 2011 SES	0.01	0.03
	(0.04)	(0.06)
6th Grade Dummy	0.68***	0.73***
	(0.04)	(0.04)
Constant	-	-0.18***
	(0.17)	(0.06)
Observations	2289	2289

Notes: Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. Child's 2011 SES is a composite measure of the child's socio-economic background as measured in 2011. The variables included in the factor analysis are the quality of the housing (floor, roof, walls, electricity), the assets (phone, motorcycle, fridge, car), and the occupation of the father – Higher values of the factor associate with higher economic status. The treatment is not correlated with the measure of SES in 2011.

Table 22: The effect of baseline socio-economic status (using district SES)

	Math	English
WSD	-0.04	0.00
	(0.08)	(0.09)
Grant	-0.08	-0.08
	(0.07)	(0.08)
District 2004 SES	0.14**	0.30***
	(0.07)	(0.09)
WSD ×District 2004 SES	0.02	-0.10
	(0.10)	(0.11)
Grant ×District 2004 SES	0.00	-0.13
	(0.10)	(0.11)
6th Grade Dummy	0.70***	0.77***
	(0.04)	(0.04)
Constant	-0.26***	-0.29***
	(0.17)	(0.06)
Observations	3659	3659
R Square	0.13	0.16

Notes: Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. District 2004 SES is the district level composite measure of the socio-economic background as measured in 2004 – Prior to the interventions. The variables included in the factor analysis are the quality of the housing (floor, roof, walls, electricity), the assets (phone, motorcycle, fridge, car, TV, fan, generator, livestock), and the expenditure on educator the past 12 months – Higher values of the factor associate with higher economic status of the district.

Table 23: The effect of baseline school level test scores

	Math	English
WSD	-0.04	-0.03
	(0.08)	(0.09)
Grant	-0.08	-0.11
	(0.07)	(0.08)
Baseline Test Score	0.14**	0.30***
	(0.07)	(0.09)
WSD × Baseline Test Score	0.02	-0.10
	(0.10)	(0.11)
Grant × Baseline Test Score	0.00	-0.13
	(0.10)	(0.11)
6th Grade Dummy	0.70***	0.77***
	(0.04)	(0.04)
Constant	-0.26***	-0.29***
	(0.17)	(0.06)
Observations	2313	2313
R Square	0.04	0.07

Notes: Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. District 2004 SES is the district level composite measure of the socio-economic background as measured in 2004 – Prior to the interventions. The variables included in the factor analysis are the quality of the housing (floor, roof, walls, electricity), the assets (phone, motorcycle, fridge, car, TV, fan, generator, livestock), and the expenditure on educator the past 12 months – Higher values of the factor associate with higher economic status of the district.

Table 24: Do wealthier district contribute more to funding the schools?

	Marginal effect of 2004 District Level SES
Gave books to school	-0.01
	(0.04)
Cash contribution	0.04
	(0.04)
Building supply	-0.03*
	(0.02)
Furniture contribution	0.00
	(0.01)
Food contribution	-0.03
	(0.04)
Observations	3659

Notes: Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. District 2004 SES is the district level composite measure of the socio-economic background as measured in 2004 – Prior to the interventions. The variables included in the factor analysis are the quality of the housing (floor, roof, walls, electricity), the assets (phone, motorcycle, fridge, car, TV, fan, generator, livestock), and the expenditure on educator the past 12 months – Higher values of the factor associate with higher economic status of the district. The coefficients are the marginal effect of the District's 2004 SES on the dependent variable.

Table 25: **Test scores before and after by intervention group**

	WSD				Control			
	3rd Grade		5th Grade		3rd Grade		5th Grade	
	2008	2010	2008	2010	2008	2010	2008	2010
Math (0-100)	32	36	59	56	35	36	59	58
	(22)	(23)	(25)	(24)	(22)	(23)	(25)	(24)
English (0-100)	35	35	48	48	34	35	47	49
	(11)	(12)	(18)	(18)	(10)	(12)	(17)	(18)
14 - 8 (% correct)	45	45	65	66	47	47	64	66
11 + 5 (% correct)	65	67	89	84	72	71	88	88
2 × 33 (% correct)	9	11	46	38	12	11	45	41
Observations	1484	1445	1359	1424	1431	1519	1367	1421

Notes: Standard deviations in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The test of comparison of mean is between years.

Figures

Figure 1: Example of drawing during the training
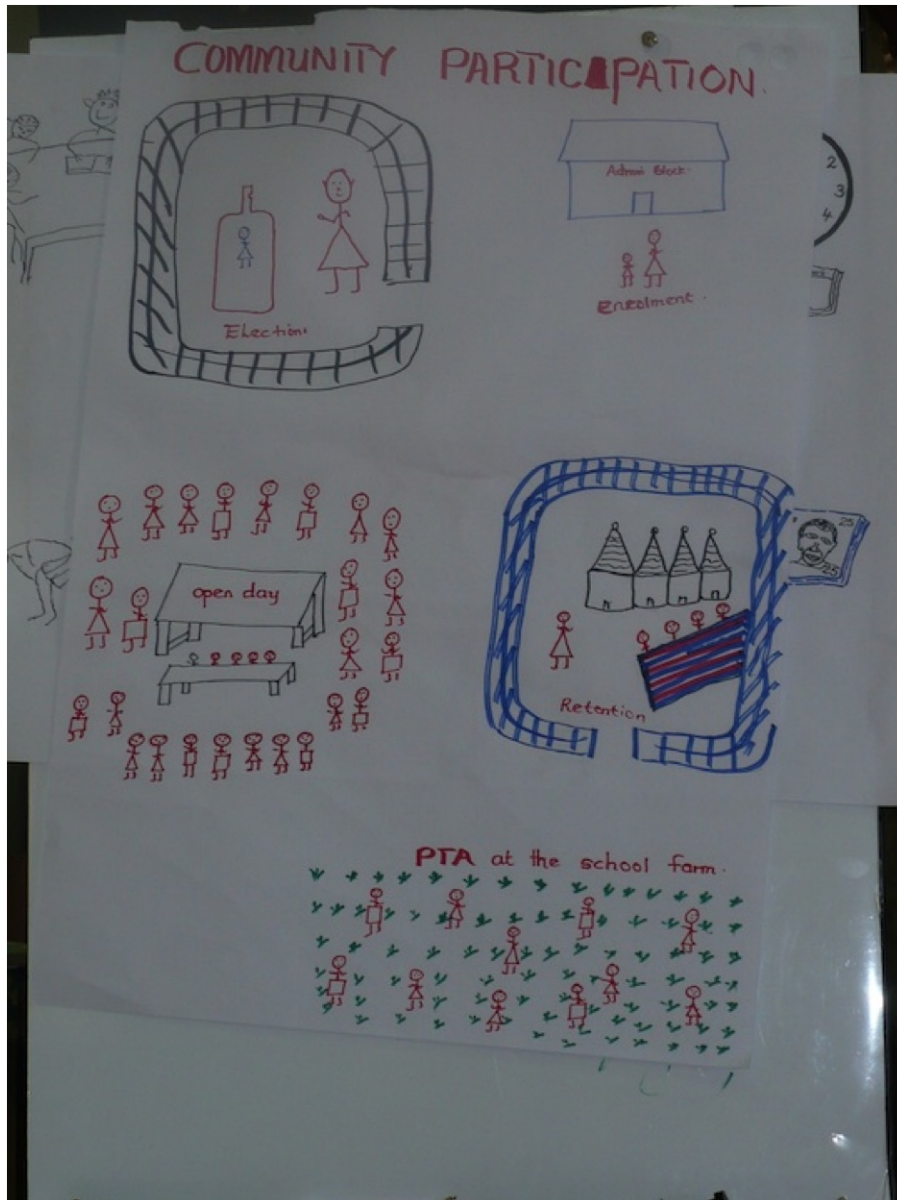
Figure 2: Geographical distribution of the schools
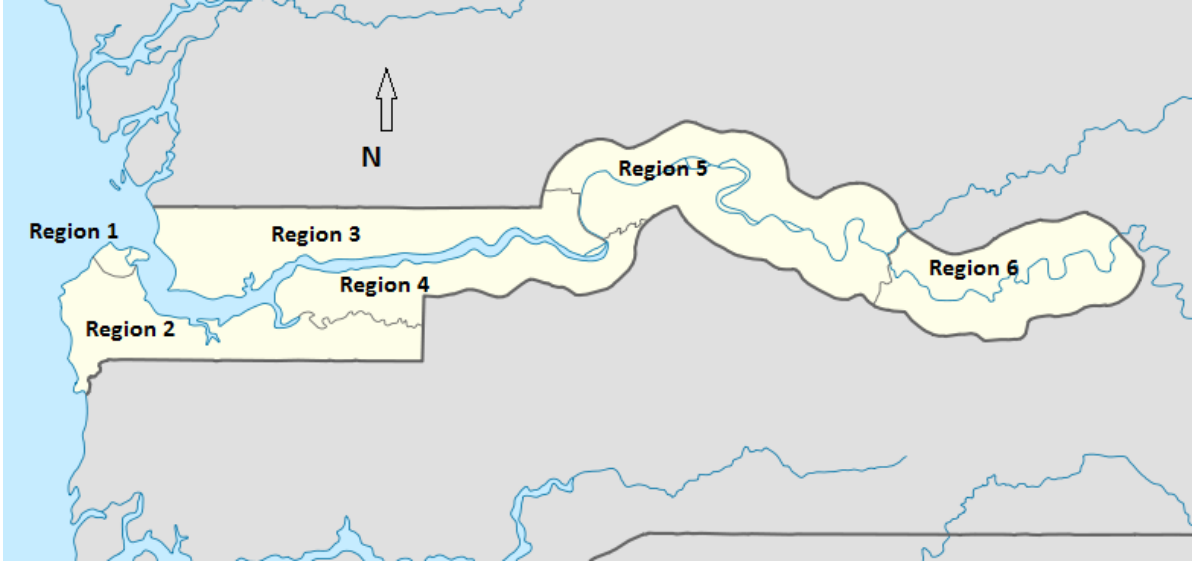
Figure 3: Sampling procedure
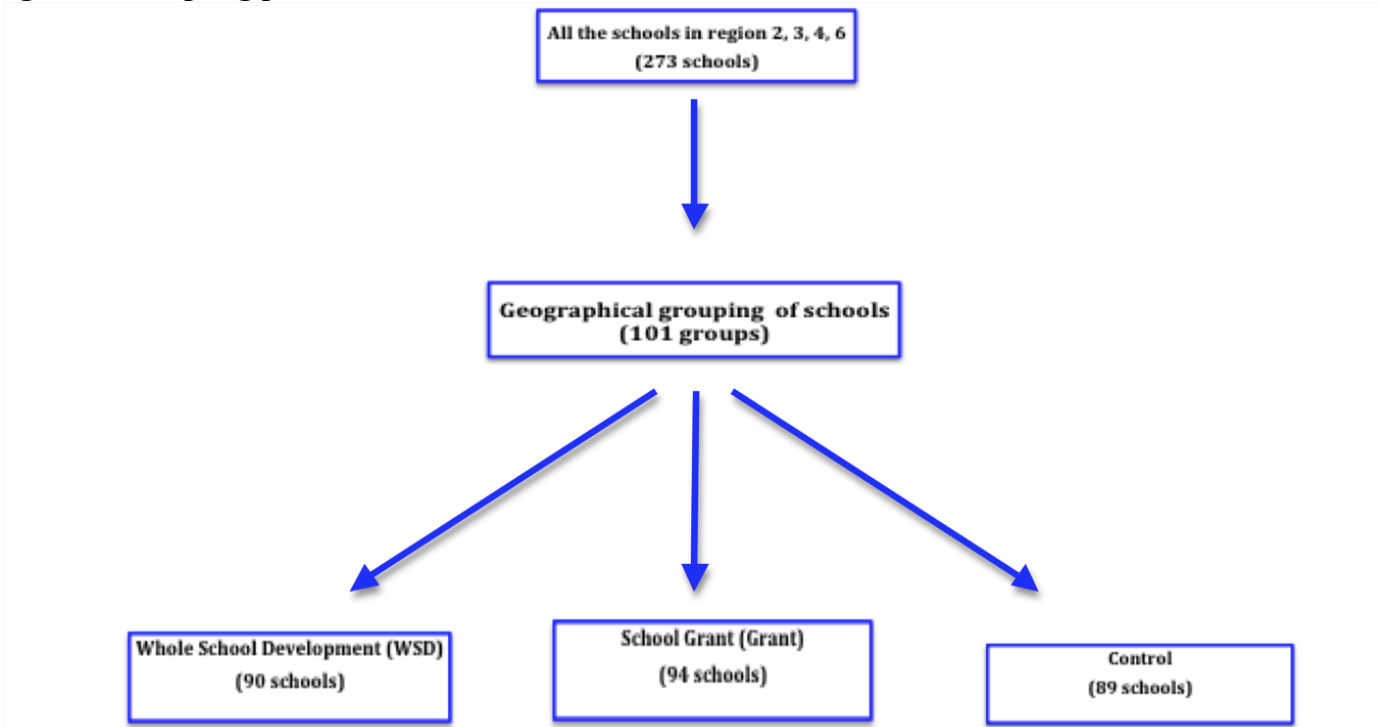
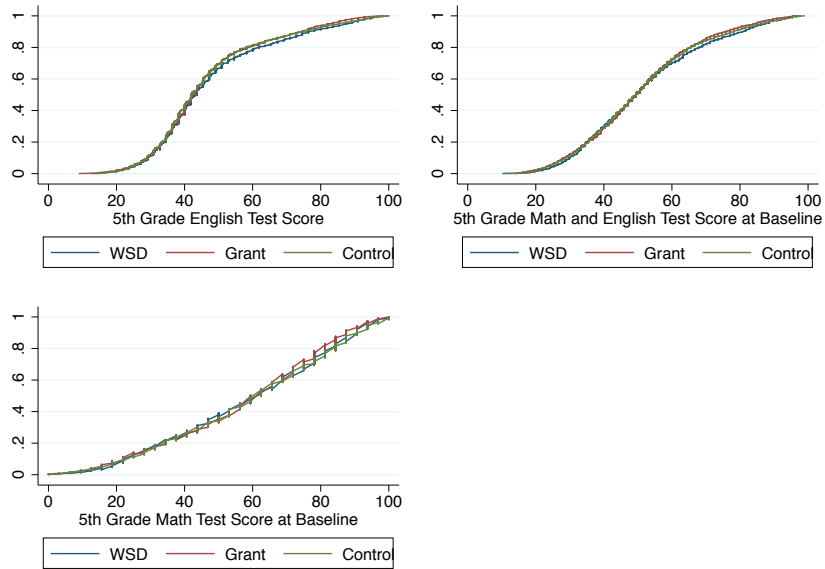**Figure 4: Fifth-grade test scores at baseline (cumulative distribution)**

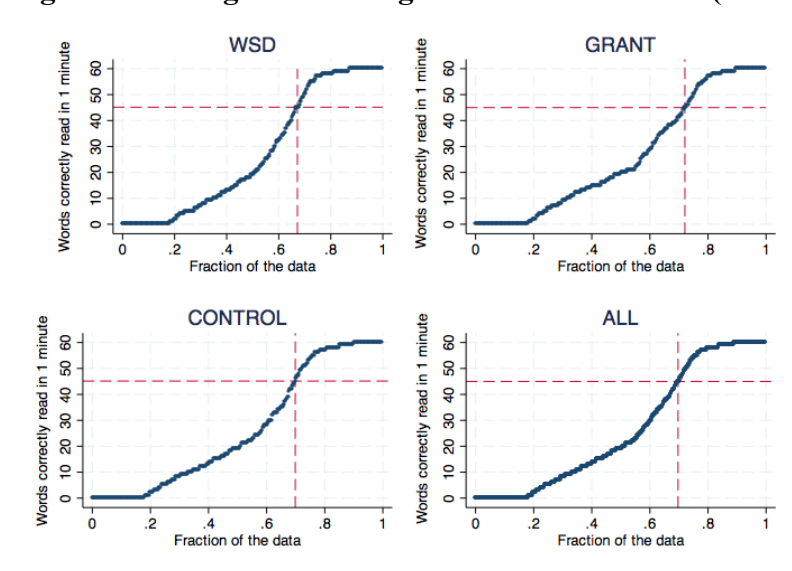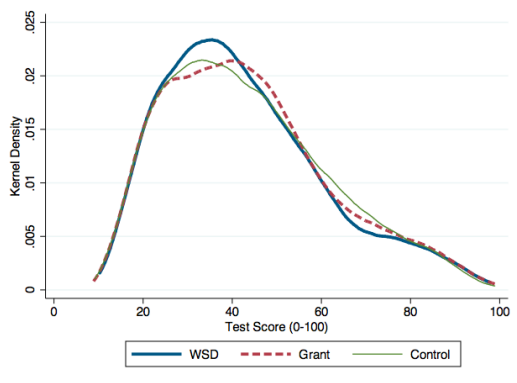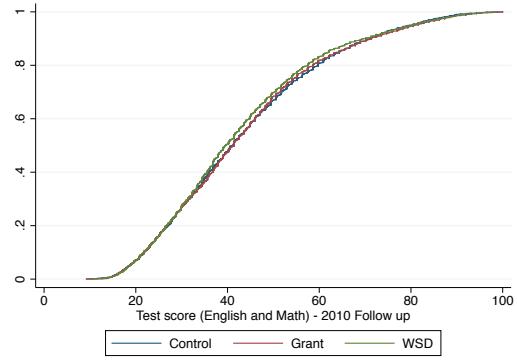**Figure 5: Fifth-grade reading outcomes at baseline (cumulative distribution)**

Figure 6: Distribution of composite test scores two years into intervention

Probability distribution	Cumulative distribution

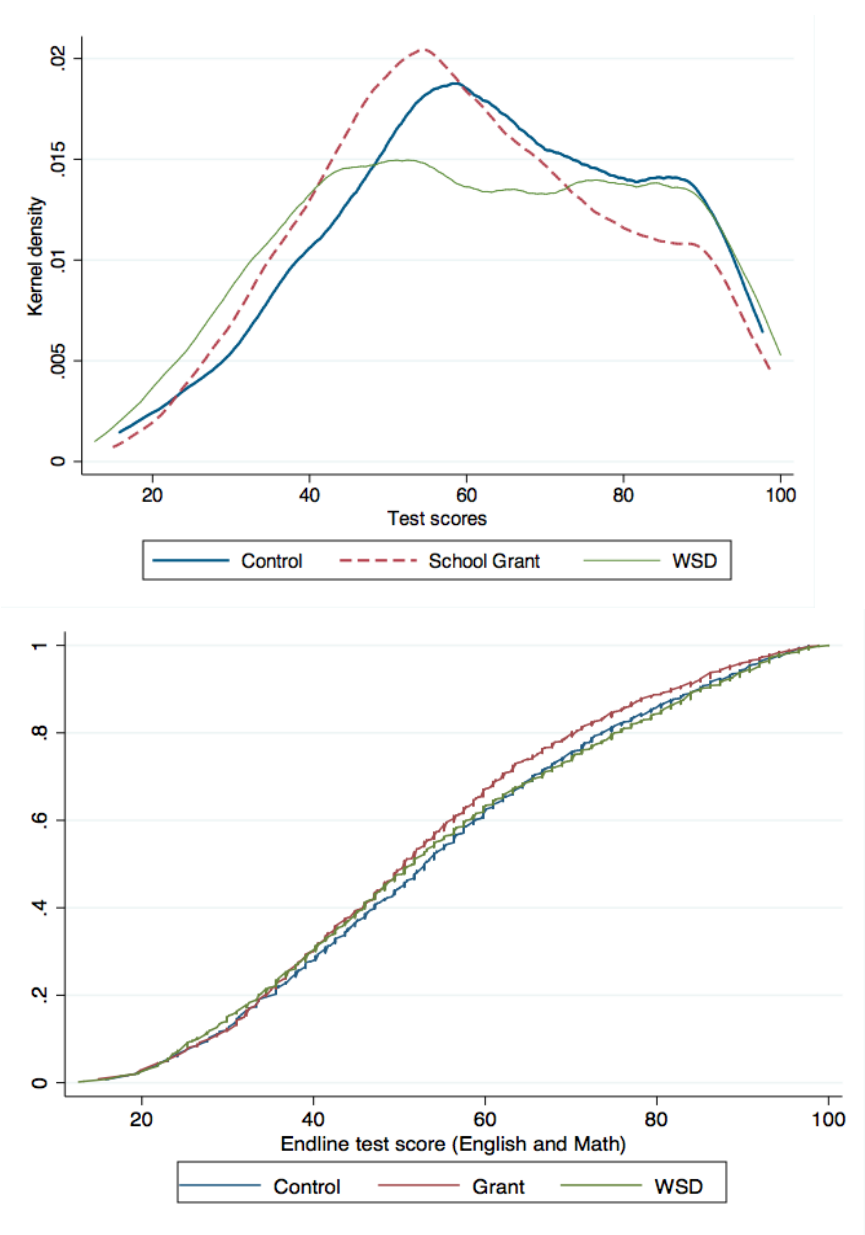**Figure 7: Distribution of composite test scores at endline (3-4 years into intervention)**

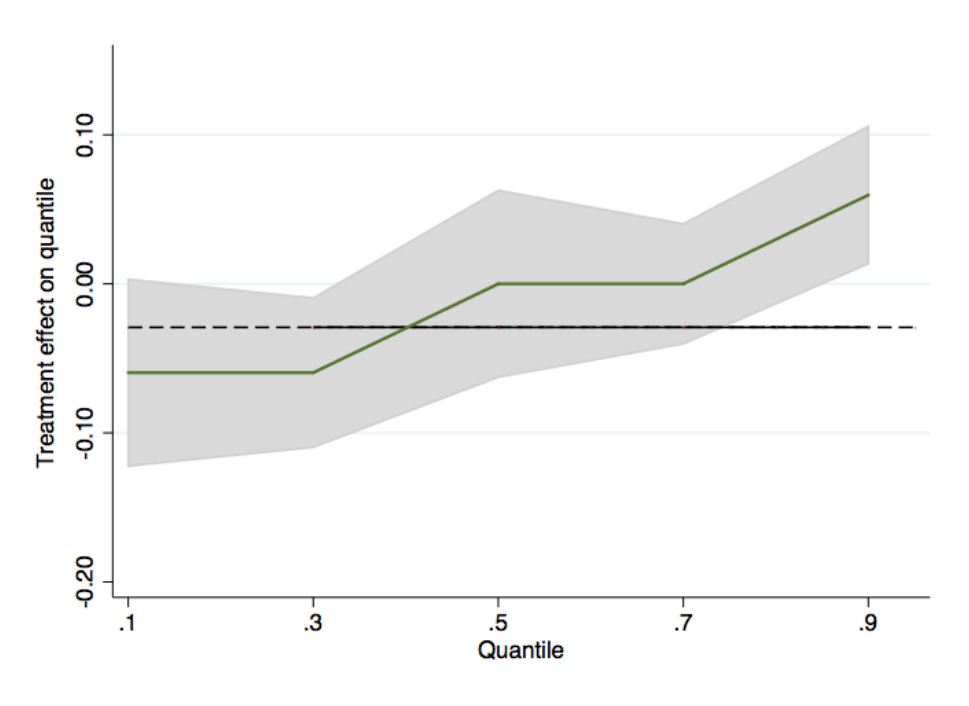**Figure 8: Treatment effect on composite student test scores by quantile**

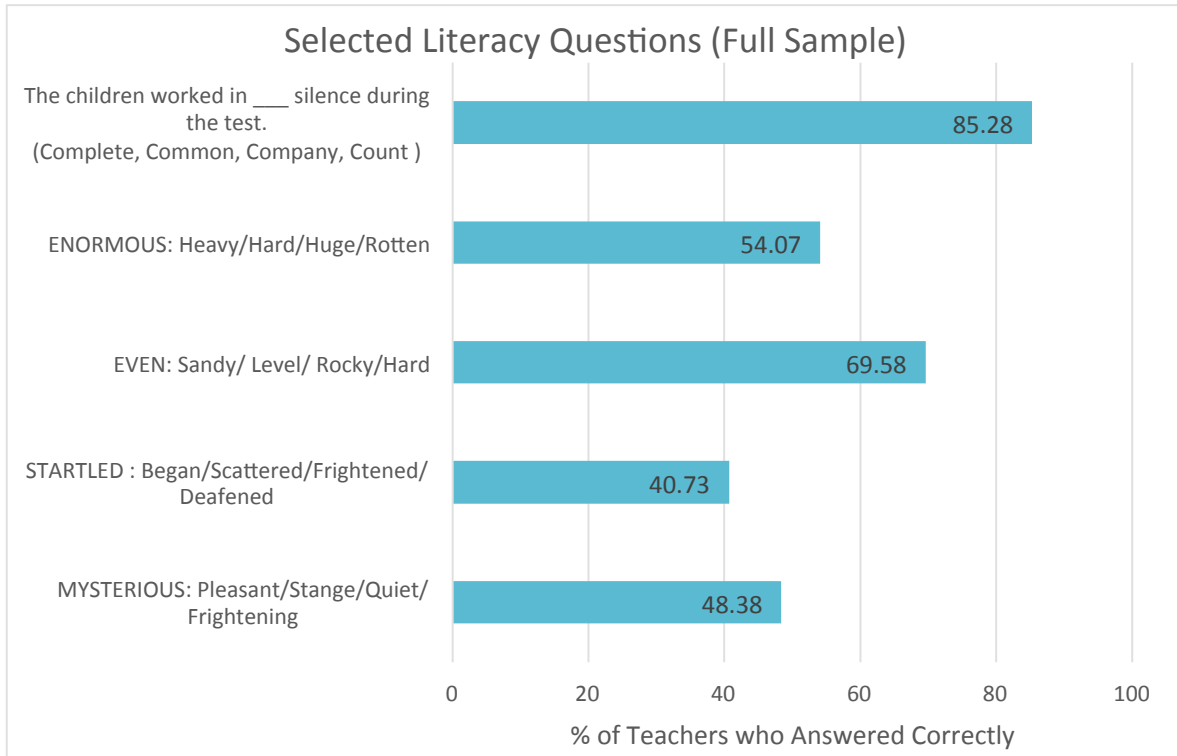**Figure 9: Teacher Content knowledge on selected English questions**

Selected Literacy Questions (Full Sample)

Question	%
The children worked in ___ silence during the test. (Complete, Common, Company, Count)	85.28
ENORMOUS: Heavy/Hard/Huge/Rotten	54.07
EVEN: Sandy/ Level/ Rocky/Hard	69.58
STARTLED : Began/Scattered/Frightened/ Deafened	40.73
MYSTERIOUS: Pleasant/Stange/Quiet/ Frightening	48.38

% of Teachers who Answered Correctly

Figure 10: Teacher content knowledge on selected math questions

Selected Math Questions (Full sample)

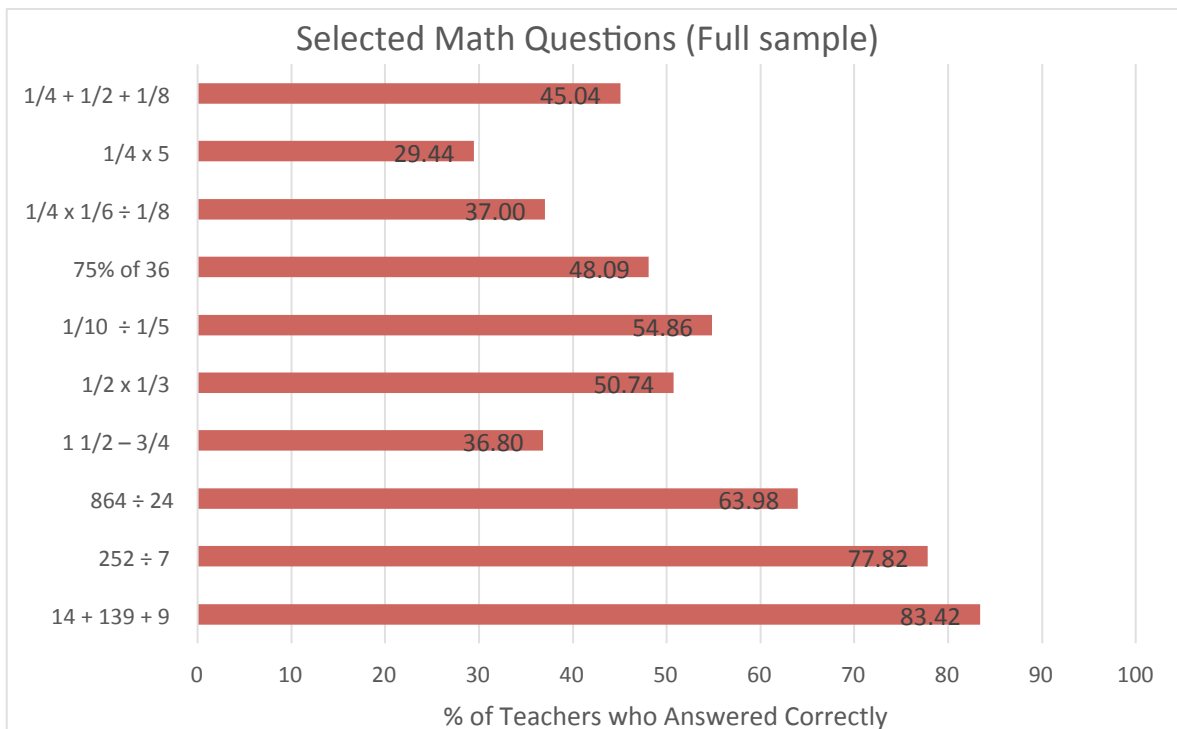| Question	%
1/4 + 1/2 + 1/8	45.04
1/4 x 5	29.44
1/4 x 1/6 ÷ 1/8	37.00
75% of 36	48.09
1/10 ÷ 1/5	54.86
1/2 x 1/3	50.74
1 1/2 − 3/4	36.80
864 ÷ 24	63.98
252 ÷ 7	77.82
14 + 139 + 9	83.42

% of Teachers who Answered Correctly

Figure 11: Correlation between teacher content knowledge and student test scores
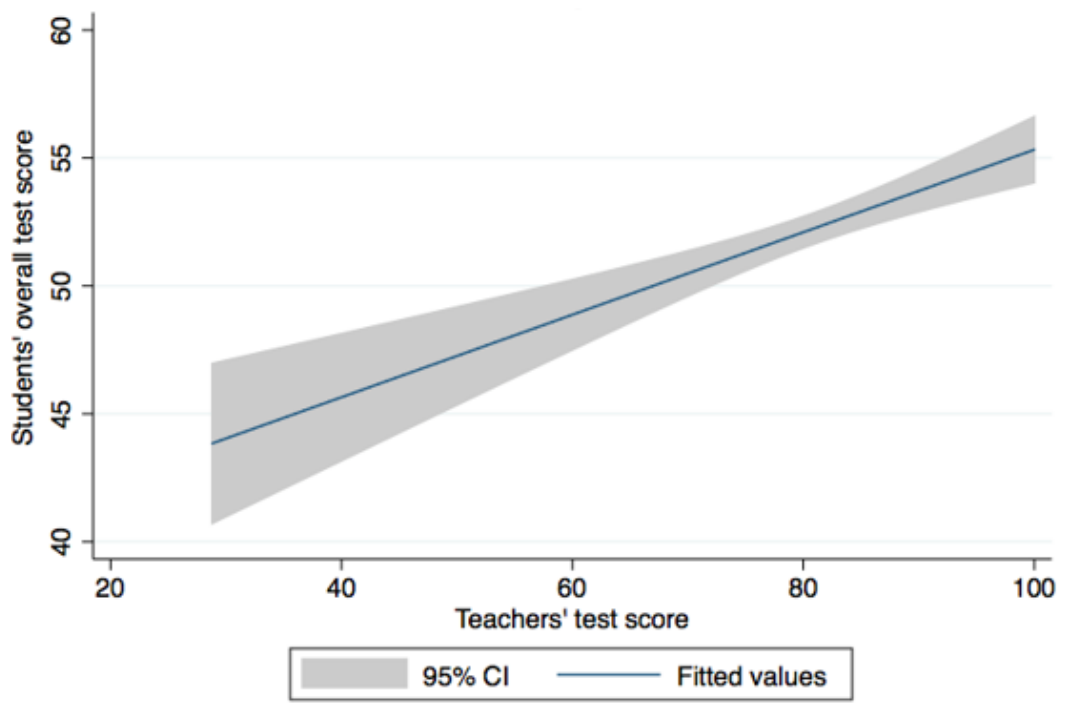

Figure 12: Level of baseline adult literacy and effectiveness of the WSD on composite student test scores
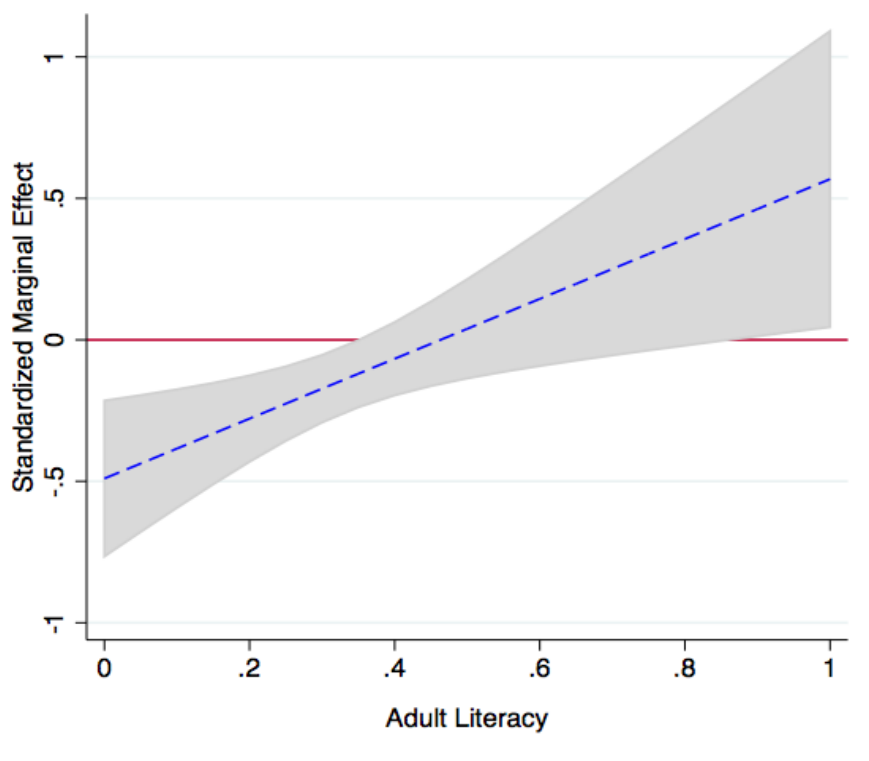

Figure 13: Level of baseline adult literacy and effectiveness of the WSD on composite student test scores: Non-parametric estimate