School-Based Management and Educational Outcomes
Lessons from a Randomized Field Experiment*

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

August 2014

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; however, low levels of human capital at the local level may offset gains from local information. This research evaluates the effectiveness of a comprehensive school-based management and capacity building program called Whole School Development (WSD). The WSD program provided a grant and a comprehensive school management-training program to principals, teachers, and representatives of the community 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 group served as a control group and received neither. We randomly assigned 273 Gambian primary schools to each of the three groups. Three to four years into the program, we find that the WSD intervention led to a 21% reduction in student absenteeism and a 23% reduction in teacher absenteeism, with no impact on learning outcomes measured by a comprehensive test. We found that the effect of the WSD program on learning outcomes is strongly mediated by the baseline local capacity measured by adult literacy. This result suggests that, in villages with high literacy, the WSD program may yield gains on students' learning outcomes. However, in villages where literacy is low, it could potentially have a negative effect. We present additional results to explore other determinants of the success of this type of intervention in low-income countries. We found no effect of receiving the grant alone relative to the control on either test scores or on participation.

JEL Classification: O15, I21, C93.

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 the delivery systems. 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). In Africa, countries

*The authors would like to thank the Ministry of Basic and Secondary Education in The Gambia for unceasing collaboration on this study. The authors also acknowledge the World Bank's Africa Program for Education Impact Evaluation, of which this study is a part, and the Education Program Development Fund for funding. Arianna Legovini and Jee-Peng Tan provided valuable guidance. Thanks to the participants of seminars at SIEPR (Stanford University), New York University, Stanford

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

like Ghana, Niger, Senegal, Madagascar, Kenya, Burkina Faso, and Mozambique have already embraced variants of this approach in their education systems.

In this research, we assess the medium run impact of this type of program in the Gambia. We address the importance of the baseline local capability on the success of school-based management policies. 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). However, local leadership or members of the community may also lack competency (relative to the central leadership) to design or implement processes necessary to tackle those problems. Therefore the effect of such a policy is ambiguous. This paper uses a large field experiment in the Gambia to evaluate and draw lessons from a comprehensive school based management program – called Whole School Development, or WSD. It is a holistic school-based management and capacity building program. This study lasted between 2007 and 2011.

In WSD schools, principals, representatives of teachers and the communities received training in a variety of areas: school leadership and management, community participation, curriculum management, teacher professional development, value and use of teaching and learning resources (e.g., textbooks and libraries), and the school environment. Through this training, the schools' stakeholders (including the community) developed a school management plan addressing the short-term 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 the schools initiate the implementation of their plan, the Ministry of Education provided a grant worth approximately 500 USD. To separate the effect of the grant from the training, another set of schools received a grant of the same size but without the comprehensive school management-training program (called Grant-only schools).

In addition, a new school constitution had been developed by the Ministry of Education as part of the School Management Manual (SMM) to enhance cooperation in schools 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 the school development that relates directly to teaching and learning (e.g., constructing teacher housing would not be an acceptable use). Finally, the control schools received neither a grant nor the management training. We assigned randomly 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 a comprehensive test in Mathematics and English. However, we found that the intervention led to a reduction in student and teacher absenteeism respectively by 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 participation. If the reduction in students' 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 on the treatment effect on test scores. Our estimates indicate that, once corrected for selection, the average treatment effect's range is -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 the 2006 average district level adult literacy. The estimates yield a positive and significant effect of the interaction term. The results remained qualitatively the same when we replace the district level adult literacy by the share of the School Management Committee (SMC) members who have no formal education (i.e., cannot read or write). Our findings suggest that 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 on 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 finding is 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 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. 2012). They found that training the community to specifically monitor teachers, combined with reduced class size and teacher incentives, yielded significant gains in various outcomes. They also found that the hiring of an additional teacher reduced the effort of existing teachers. However, in the intervention where the communities were involved in monitoring, the negative impact on teachers' effort dropped significantly, leading to improvement in learning outcomes. Our finding is also in contrast with 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 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; Reinikka and Svensson 2004). However, we found 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 School Management Committee (SMC) and the regional directorate. The schools were required to subsequently submit the receipts to the regional directorate. In addition, there are officials at the

regional directorate, called "Cluster monitors" whose role is 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, are at play.

Finally, even in an environment where local capture is limited or controlled, the capacity at the local level 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, the conventional wisdom seems to support 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 contexts of high and low-income countries, and even between India and countries like Gambia, renders 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 inadequate other educational inputs are the main factors explaining the limited impact of the intervention.

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 six first 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 has 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) carries specific information about Gambian schools (Adebimpe, Blimpo, and Evans, 2009). We found that overall the basic infrastructure of schools is in good condition.[3] The main buildings (classrooms and staff headquarters) are overall in good condition throughout the four regions. Of the 273 schools visited, 9% require some minor repairs for the walls,

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

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 are 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; however, 65% of PTAs have no funding. Head teachers were asked to report the most important challenge that the school faces 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. In the region with the greatest number of teachers with a lesson plan, only 62% 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.

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,[4] and 69% owned a bicycle.

[4] Either the landline or a person in the household who possesses a mobile phone.

3. Experimental design

3.1. The interventions

The main intervention evaluated in this paper is a holistic school management capacity building program called Whole School Development (WSD). 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 the training. We compare these two interventions to a control group that received neither the grant nor the training. Table 1 provides a snapshot of the key element of the interventions and Table 2 provides a brief summary of the timeline.

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 of what kind of data and information are relevant and should be collected on a regular basis (e.g., monitoring learning outcomes and absenteeism). It emphasizes how to analyze the data and plan for short-term and long-term solutions to school problems. The second step is the implementation of the resulting plan. Finally, 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 Management Training

The management training and capacity building are the centerpiece of the WSD intervention. The principals, teachers, and representatives of parents and students receive training in a variety of areas presented in the school management manual. In the course of this training, participants develop a local school development plan addressing various areas with guidance from of the trainers and the supervision of the WSD unit within the Ministry of Education. The training took place in three steps. In the first step, the experts who developed the SMM trained twenty people at the national level. Those twenty people were called ''master trainers''. They were dispatched simultaneously in teams of five to the four regions. In the second step, the master trainers trained people at the regional level. And finally, the people trained at the regional level trained the representatives of the teachers and the community. Since most parents do not speak, read, or write English, the training put some emphasis on the local languages and drawings (See Figure 3) to convey the messages more effectively. Throughout this process, our research team together with BESPOR played a monitoring and supervisory role.

The Grant

Some of the activities suggested in the manual and included in the school development plans, like workshops, might require some funds. 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 $500 was given to

all the schools in the WSD and the Grant-only groups. 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. The two regions that were excluded from the study were Region 1, which is essentially 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 1 summarizes the sampling procedure and Figure 2 shows the geographical distribution of the schools by intervention group. 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.[5] The randomization was further stratified by school size and accessibility.[6] We will discuss the effectiveness of the randomization in detail later, but each group proved to be similar at baseline. 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.

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 the disbursement of the grants was not complete and many schools that had received their grant had not yet used it. This information was obtained from the regional directorates who were the key intermediaries for the grant disbursement process. This problem of slow disbursement of education grants by local committees was also observed in western Kenya (Conn et al. 2008) and in Niger (World Bank 2011).

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 school and the students, conducted classroom observation, and gave a literacy and numeracy test.[7] Unless otherwise indicated, the following data were collected at each of the four rounds of data collection; Table 3 provides detailed information.

---

[5] 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 he is assigned to.

[6] The Ministry defines accessibility through "hardship status". Schools that are most remote receive an allowance from the Government, as discussed in Pugatch & Schroeder (2013).

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

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 was about 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.

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.

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 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.

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 2010 in all the schools whenever possible.[8] Students for the pupil interview were selected randomly from among those who participated in the written test. At baseline, we interviewed 2,696 students in total: 879 from WSD schools, 920 from Grant-only, and 897 from the control schools.

Teacher content knowledge

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

Qualitative data

In 2010, we added many open-ended questions to the head teacher interviews to collect some information about their visions regarding school management. We also addressed similar questions to a few households whose children are in the relevant schools. We have also been heavily involved on the ground for the entire first year of this program. Our various conversations with the government, the schools, and the communities add important information that is useful for a better understanding and rationalizing of the findings.

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

5. Identification, empirical strategy, intermediate outcomes

5.1. Identification and group comparison

In a 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 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. The WSD program 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. The difference is significant. More WSD schools reported also having received financial or in-kind support from the community, though the difference is not statistically significant. This probably reflects information about the program that preceded program implementation and the survey. While one might that higher contributions indicate greater affluence among WSD schools, we find that WSD schools are not higher in their access to a tap for drinking water (23% of WSD schools versus 20% of Grant-only schools and 33% of control schools).

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 should not be systematic differences among the three groups in unobserved characteristics either.

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, $GRANT_s$ = 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.

One important aspect of the WSD is community participation. We observe that the first grade enrollment was about 16 students higher on average in the WSD schools relative to the control schools, from a base of about 70, although this difference is not statistically significant (Table 7). All schools in both groups reported having a PTA: However, over 70% of them reported having no funding. Fundraisers and member contributions remain weak. The WSD group (46%) reported having received support in cash and in-kind from the community more than the control group (35%), although this difference is also not significant at standard levels (p-value = 0.15).

Over 65% of the schools have a staff code of conduct in both groups. Although the control group reported more teacher mentoring systems (6% more and statistically insignificant), there are more trained mentors (14% more and statistically significant) in the WSD group relative to the control group. Written school policies were infrequently observed in either group, but the WSD group had more often developed written policies than the control (45% in WSD and 36% in the control, but not significantly different; results reported).

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.[9] There is a higher rate of establishment of various school management committees (SMC) in WSD schools, as recommended by the School Management Manual (SMM). 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 rather the 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 teacher preparedness, the control schools appeared to perform better one year into the program (Table 8). We observed teachers' written lesson notes for the day of the visit in more control

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

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.[10]

Overall, pupil participation in terms of asking questions to teachers is poor. It is roughly the same in WSD classrooms (26%) and the control classrooms (23%) (insignificant difference). The use of textbooks during the visit was more frequent in the control group (47%) than the WSD group (38%), a marginally significant result. However, the workbooks were used more in the WSD group (54%) than in the control group (45%) (not significant).

Absenteeism remains pervasive. 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 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 8).

We found no difference between the two groups in terms of student performance (Table 7). 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 intent-to-treat 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 other reasons.

Even though we found 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. 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 11).

[10] 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.

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 IV 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 VII, 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. Even if some of the changes observed may be expected to impact student learning, we found 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 both teachers and 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.[11] We run the same model where the outcome variables are student absenteeism and teacher absenteeism. The estimates in the first column of Table 14 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 is the same model but with teacher absenteeism as the outcome variable. The WSD reduced teacher absenteeism by about 3-percentage point from a base of about 13%, which represents about a 23% reduction in teacher absenteeism.

[11] 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.

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 left shift of the distribution of test scores, albeit only at the left tail. This is suggestive evidence for the hypothesis that the WSD program attracted more low-performing students. Miguel and Kremer's (2004) de-worming intervention in Kenya found large effect on participation, but found no effect on test score; this 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 poorer, then one would 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 quantiles. Figure 8 shows an upward trend, which partially supports this story. For this effect however 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.) This is clearly a strong assumption. We address this selection issue by bounding the treatment effect using Lee's trimming procedure (Lee 2009). The procedure consists of trimming out a proportion of the lower tail (respectively upper tail) of the distribution in the WSD group order to construct an upper bound (respectively lower bound) of the effect of the intervention. 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 the sample counterpart of the following:

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

and

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

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 on 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 treatment (WSD). 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 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 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 might not be pronounced. 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.

6.2.2. Poor complementary inputs: Teacher quality

A third explanation is that other inputs such as teacher quality are sufficiently low that increased participation will not necessarily 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. Figures 12 and 13 sample questions and teacher performance on them. The findings suggested that teacher content knowledge was problematic. Only 2.6% of teachers scored 95% or more, and over one-third of the teachers scored below 75%.

Figure 14 shows a positive correlation between matched teacher and pupil test scores. 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 19), to be contrasted with estimates between 52% and 65% in a sample of Latin American countries (Bruns et al. 2014). Taken together, these results suggest that teacher quality and effectiveness may be so low in Gambia so that other school improvement interventions will not work.

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. It is reasonable to assume that some level of human capital is needed at the local level for interventions such as the WSD to anchor on: In other words, 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.

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

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

We report estimates of equation 2 in Table 17, where $BaselineHC_d$ is the district level adult literacy in 2006. Across the districts in 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 true 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 sufficiently low. The channel for this potential negative effect could come from the cost of shifting from one set of management practices that are functioning to some degree to another set of practices that are presumably better. If the new practices 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 18 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 9 and 10. We conclude that the WSD intervention is likely to improve learning outcomes in area with high baseline human capital, but it could be counter-productive 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 was 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%) disapprove of this idea, 19% think that it would be 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. The following two responses are typical of the arguments put forth:

- "Yes, because there is more interaction between the teachers, pupils and parents…. So the school and community know better [and] how to manage the affairs of the school.''

- "Yes, because they are the [ones] on the ground, who knows what is good for them, and make their administrative work easier.''

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.

- "Very good. Teachers know that their hiring and firing are in the hands of the SMC [community] effective teaching and learning will take places. They will all be serious at work.''

These are legitimate arguments to support the position of those in favor of a decentralized management. However, most (over 75%) 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. The following selected quotes are representative of the modal responses:

- "It will not be good to give all the power of decision to the community in managing the curriculum, for some communities are not educated western-wise and in this they cannot manage a school.''
- "It will not be a good idea, as there will not be a fair play and the know-how will also be lacking among the community.''

- "No it is not a good thing. A greater percentage of the communities where most of the schools are located cannot read and write."
- "Almost impossible because a large portion of our communities are illiterate."

Others believe that it could bring conflict within the communities. As one teacher puts it:

- "No! If such powers are given to the community they could abused it and could even bring conflict in the school."

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 highest level and hold high profile position such as doctors, ministers, etc. Therefore, it appears that the parents care about the educational outcomes of their children, but there is a contrast between this aspiration with 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. These assessments indicate that those parents may be more educated and better able to assess the progress of the children and the performance of the school.

These findings confirm that the WSD may be more appropriate where local capacity is sufficiently high. Tables 20a and 20b interacts the treatment with the baseline socio economic status of the students and their district. The interaction effect is in signification. In Table 20c we further present evidence that higher economic status does not associate with higher likelihood of financial or in-kind contribution to the school. We interpret this as evidence that the human capital effect is not a proxy for income.

6.2.4. Improvement in the control schools?

The lack of impact on test scores could be also due 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 an orthogonal set of practices from the WSD schools without any support. However, 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 the control group (as well as in the other groups). Table 15 presents the results of such analysis. 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 a comprehensive test. However, we found a large effect on participation: The intervention led to reductions in student and teacher absenteeism respectively by 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 the 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. We also found 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 the inability of the parents to 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 understand the performance of their children and thus to demand for accountability from educators. That is precisely what the capacity building component of the WSD attempted to address. The WSD does not appear to have accomplished this. The challenge might be more with the basic inability of parents to read and write and less with the real steps that they need to take. The WSD has focused more on the latter.

We found no evidence of positive effects on outcomes due to the Grant-only intervention, only on process variables such as community engagement on decision making. However, there are many reasons why this should be taken with caution. First, the 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 for us to expect a substantial effect three years later (although note that no effect observed at any point). With an increased amount and or with more sustained yearly grants the results might differ.

Based on this study, we draw the following conclusions and policy implications. First, a structural feature that matters for an effective local management program, such as the one envisioned and studied here, is local baseline basic human capital such as literacy in the communities. We hypothesize that in general, the gap between local capacity at the central level and the local level is a key determinant of the success of this kind of policies. In countries where this gap is small, regardless of the levels, a decentralized policy would be superior because of the added value of localized information. However, if the gap is sufficiently high in favor of the central government, then the localized information is less useful because the communities are not well equipped to act on them. Our findings show that the Gambia may belong to the latter group. Other studies should explore further this dimension and an ongoing meta-analysis of other similar studies in low-income countries could shed more light on this hypothesis.

Second, in the Gambia, there appear to be other binding constraints on the education production function. Two of those constraints, explored here, are teacher capacity and effectiveness; others are limited instructional time due to the widespread double-shift schools and teacher compensation. It might

be desirable and more pressing to address these dimensions independently. Many of these constraints are better addressed by national policy.

Third, our findings suggest that a mechanism to supply accurate information to the communities (about the relative performance of their children and the schools) could be desirable. There are good reasons to believe that if well informed, parents will try to hold the schools accountable for their children's learning outcomes (among other things). This is particularly relevant for the Gambian context wherein we found that most parents, including in the rural areas, have high aspirations for their children's professional future and educational achievements. However, we also found a sharp inability of the parents to understand the performance of their children and the functioning of the schools, even after the intervention.

Our findings call for caution with this type of policy. 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. Many other countries have adopted the approach outright. These other studies will shed more light on the issue.

References

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.

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

Bardhan, Pranab. Decentralization of governance and development. *Journal of Economic Perspectives*, 16:185–205, 2002.

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.

Bardhan, Pranab and Dilip Mookherjee. Decentralizing antipoverty program delivery in developing countries. *Journal of Public Economics*, 89:675–704, 2005.

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, 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 and John Van Reenen. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, CXXII, 2007.

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

Bruns, Barbara… Building Better Teachers 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.

Conning, Jonathan and Kevane Michael. Community-based targeting mechanisms for social safety nets: A critical review. *World Development*, 30:375–394, 2002.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. *NBER Working Paper 17939*, 2012.

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.

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.

King, Elizabeth M. and Peter F. Orazem. Evaluating education reforms: Four cases in developing countries. *World Bank Econ Review*, 13:409–413, 1999.

Kremer, Michael and Alaka Holla. Improving education in the developing world : What have we learned from randomized evaluations? *Annual Review of Economics*, 1:513–542, 2008.

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.

Muralidharan, Karthik and Venkatesh Sundararaman. The impact of diagnostic feedback to teachers on student learning: experimental evidence from India. *Economic Journal*, 120:187–203, 2010.

Pugatch, Todd and Elizabeth Schroeder. Incentives for Teacher Relocation: Evidence from the Gambian Hardship Allowance. *IZA Discussion Paper* 7723, 2013.

Rai, Ashok S. Targeting the poor using community information. *Journal of Development Economics*, 69:71–83, 2002.

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

Stiglitz, Joseph. Participation and development: Perspectives from the comprehensive development paradigm. *Review of Development Economics*, 6:163–182, 2002.

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.

World Bank. School grants as catalyst for school-based management and accountability? Results from a randomized impact evaluation in Niger. *Unpublished Draft Report for the Ministry of Education of Niger* (MEN), 2011.

Tables

Table 1: Snapshot of the intervention groups

	Grant	Management Training
WSD	Yes	Yes
Grant	Yes	No
Control	No	No

Table 2: Summary of the timeline

Date	Activities
10/07 04/08	Sensitization and coordination between stakeholders
04/2008 - 06/2007	Assignment to interventions and baseline data collection
05/2008 - 12/2008	Grant distribution and training in the WSD schools
05 - 06/2009	Collection of the first follow-up data
05 - 06/2010	Collection of the second follow-up data
05 - 06/2011	Collection of the third follow-up data
Throughout	Monitoring

Table 3: Description of the data

Year	Name	Respondent(s)	# Obs.	Notes
2008	School Data	Principal, deputy	273	
	Student test	3rd, 5th grades	8856	
	Classroom Visit	4th, 6th grade	528	
	Student interview	3rd, 5th grade	2688	Subset of tested students
2009	School Data	Principal, deputy	176	No data in Grant schools
	Student test	4th, 6th grades	5660	
	Classroom Visit	3rd, 5th grades	346	
	Student interview	4th, 6th grades	1755	
	Teacher test	About 6 teachers	1049	
2010	School Data	Principal, deputy	276	
	Student test	3rd, 5th grades	9022	
	Classroom Visit	4th, 6th grades	502	
	Student interview	3rd, 5th grades	2678	
	Parents' interview	Parent or care-giver	567	Of two interviewed student
2011	School Data	Principal, deputy	274	
	Student test	4th, 6th grades	5230	
	Classroom Visit	3rd, 5th grades	534	
	Student interview	4th, 6th grades	2579	
	SMC interview	Committee less principal	249	Mostly PTAs, in controls and
	Teacher interview	4th, 6th grade teachers	517	Teachers of tested students

Table 4: Baseline Group Comparison on School Characteristics

	WSD	Grant	Control
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/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%
	(.015)	(.014)	(.012)
Primary Education or more (2003 Census)	57%	55%	55%
	(.017)	(.016)	(.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)
Percent 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

Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The test of comparison of mean is between each treatment group and 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

Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The test of comparison of mean is between each treatment group and 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		

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

Table 7: Student performance (First follow-up in 2009)

Reading test	4th Grade			6th Grade		
	WSD	Control	P-value	WSD	Control	P-value
Correct letters per	55	57	0.26	73	75	0.17
	(1.23)	(1.23)		(1.15)	(1.1)	
Correct words per	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		

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 8: 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	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		

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

Table 9: Student performance (Second follow-up in 2010)

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

Standard deviations in parentheses. *** 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, [b]Percentage of student who can read 45 or more word per minute.

Table 10: Teaching practices (Second follow-up in 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
	(I)	(II)	(III)	(IV)
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%

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

Table 11: Participation in management (Second follow-up, 2010)

	Marginal Probability to participate in decision-making						
	Teachers (I)	Parents (II)	Rely on SDP (III)	RED (IV)	Fundraisers (V)	Know PTA memb. rule (VI)	# Meetings parent/school (VII)
	Probit	Probit	Probit	Probit	Probit	Probit	OLS
WSD	0.42*** (0.08)	0.64*** (0.06)	0.18*** (0.07)	0.26*** (0.08)	0.11** (0.06)	-0.15** (0.08)	-0.41*** (0.18)
Grant	0.37*** (0.08)	0.65*** (0.06)	0.16** (0.07)	0.37*** (0.08)	0.07 (0.06)	-0.04 (0.08)	-0.26 (0.18)
Observations	274	274	274	274	274	505	505
Mean of dent variable in comparison group[a]	3.3%	9%	1%	2%	7%	50%	1.9

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 -- Intent-to-treat (2010)

	3rd Grade		5th Grade	
	Standardized test score	Probability that a child can read 45 or more word per minute	Standardized test score	Percentage of student who can read 45 or more word 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%[a]	11%[b]	52.06%[a]	38%[b]

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

[b]Percentage of student who can read 45 or more words per minute

Table 13: Average Treatment Effect on 4th and 6th-Graders exposed to the intervention continuously over 3 to 4 years

	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	0.78	0.32
Observations	4817	4817

Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. The test of comparison of mean is between each treatment group and the control group.

Table 14: Effect of the Interventions on Student and Teacher Participation

	Absenteeism		Log First-Grade Enrollment
	Students	Teachers	
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

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 a 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 of observation in columns 2-3 is the school.

Table 15: 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

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

Table 16: Bounds for the ATE accounting for selection using the trimming procedure

			Math	English
Control	Number of Observations	444		
	Proportion Non-missing	70.95 %		
	Mean (std) test score		73%	61%
			(20)	(18)
Treatment	Number of Observations	453		
	Proportion Non-missing	79.25		
	Mean (std) test score		71.12%	62%
			(23)	(21)
Lee's upper bound			0.17	0.26
			(0.06)	(0.07)
Lee's lower bound			-0.19	-0.16
			(0.09)	(0.11)

Robust Standard errors in parentheses. The dependent variable is a standardized test score.

Table 17: 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

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 18: 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 Literacy	-0.65**	-0.57*
	(0.29)	(0.34)
Grant × SMC Literacy	-0.36	-0.39
	(0.30)	(0.39)
Constant	0.41	0.64
	(0.17)	(0.21)
Observations	2035	2035

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

Table 19: 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
Based on ten two-minute snapshots of classroom activities in 534 classroom observations.				

Table 20a: The effect of baseline economic status.

	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	-0.14***	-0.18***
	(0.17)	(0.06)
Observations	2289	2289

Robust Standard errors in parentheses. *** 1% Significance Level, **5% Significance Level, *10% Significance Level. Child's 2011 SES is 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 20b: The effect of baseline economic status.

	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)
6^{th} 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

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 20c: Do wealthier district contribute more to funding the schools?

Dependent Variables	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

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

Figures

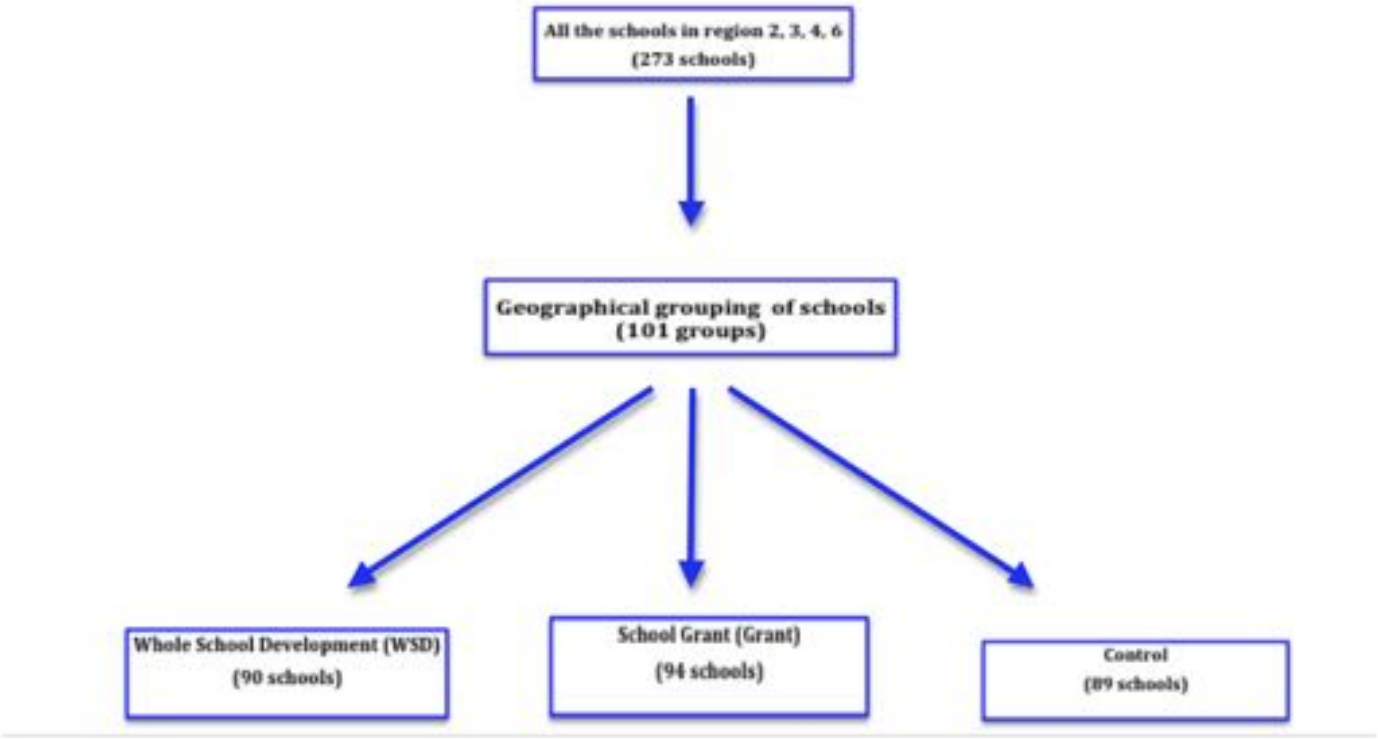

Figure 1: Sampling procedure

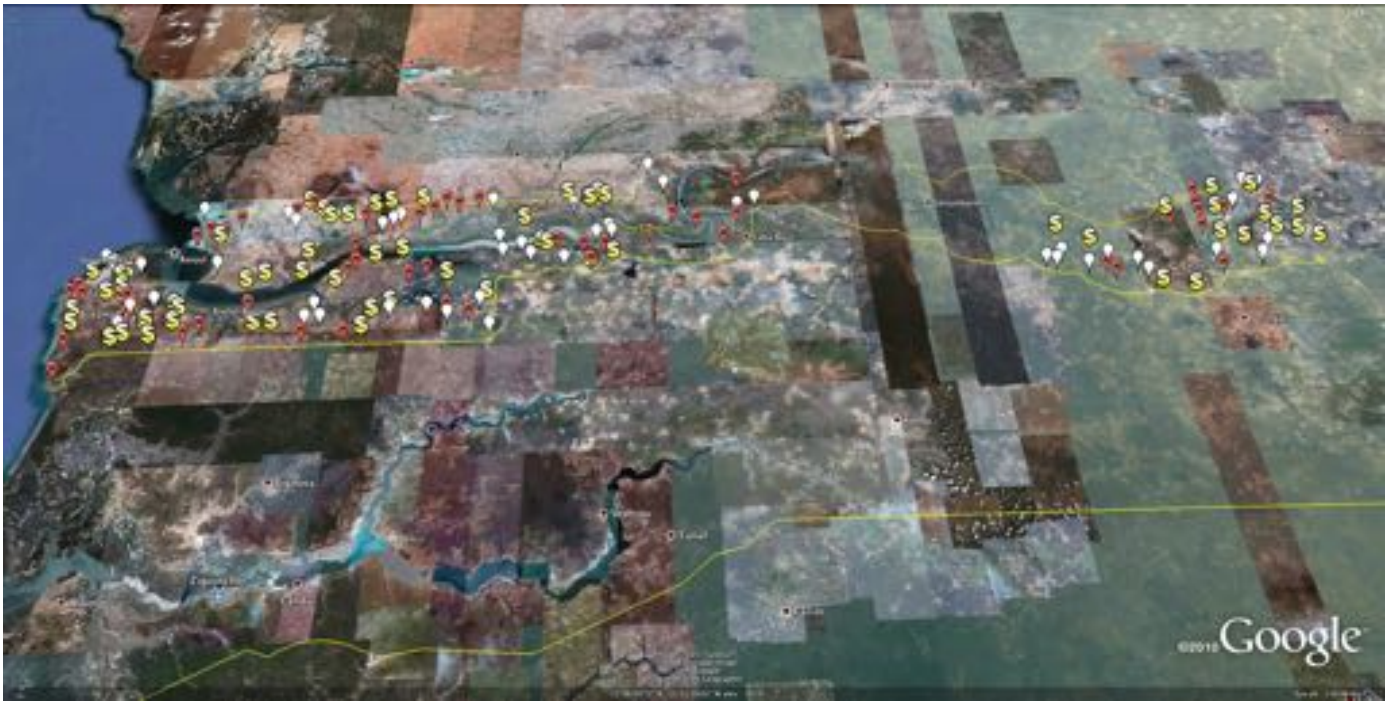

Figure 2: Geographical distribution of the schools

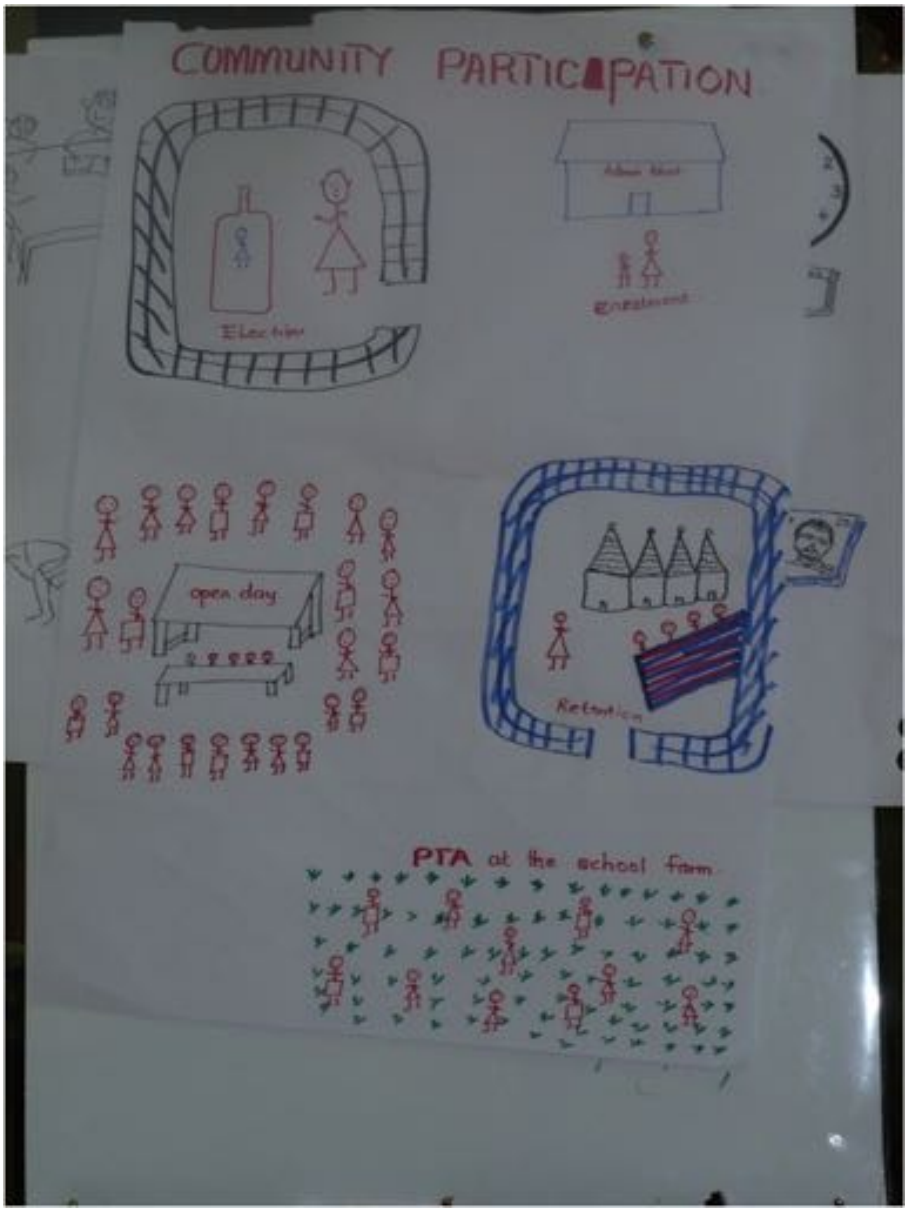

Figure 3: Example of drawing during the training

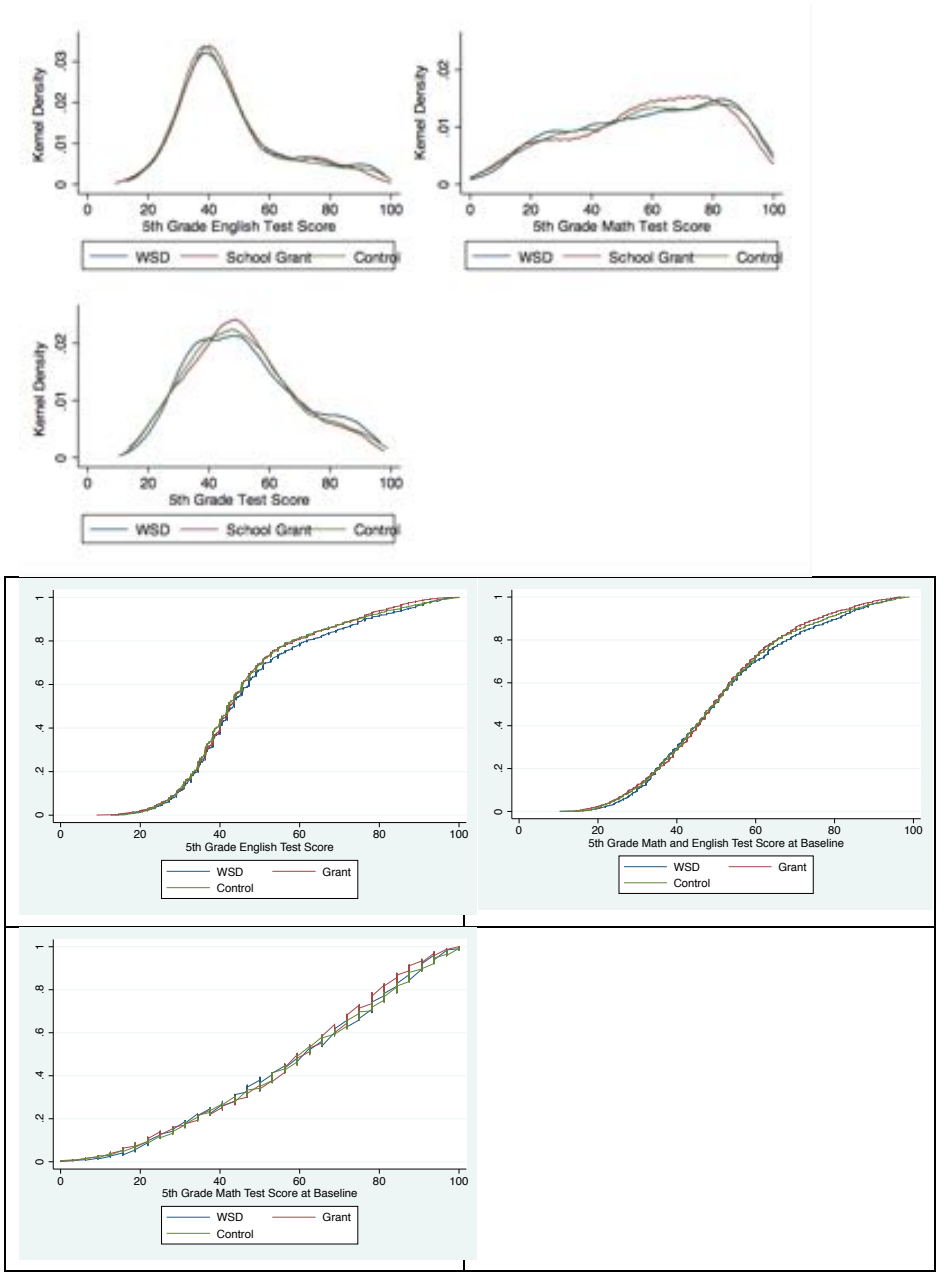

Figure 4: Fifth-grade test scores at baseline

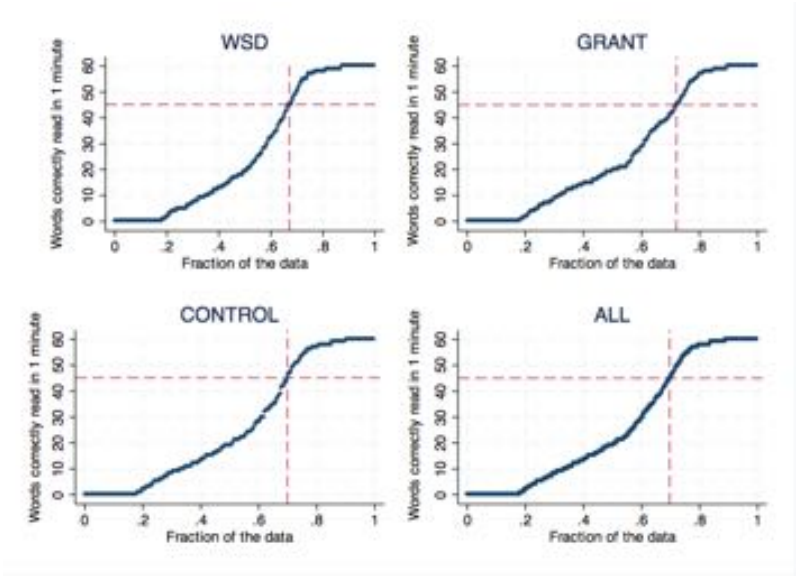

Figure 5: Fifth-grade reading outcomes at the baseline

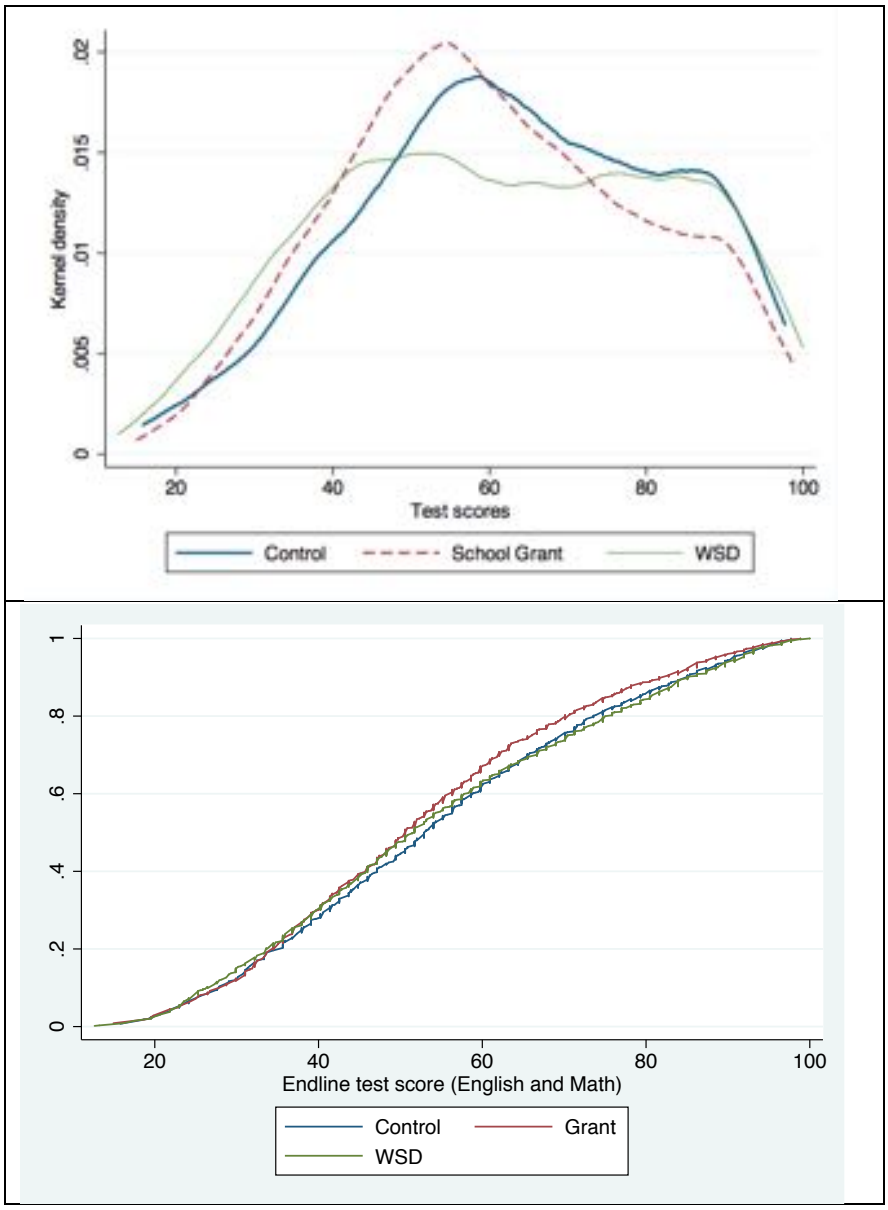

Figure 7: Distribution of composite test scores at endline

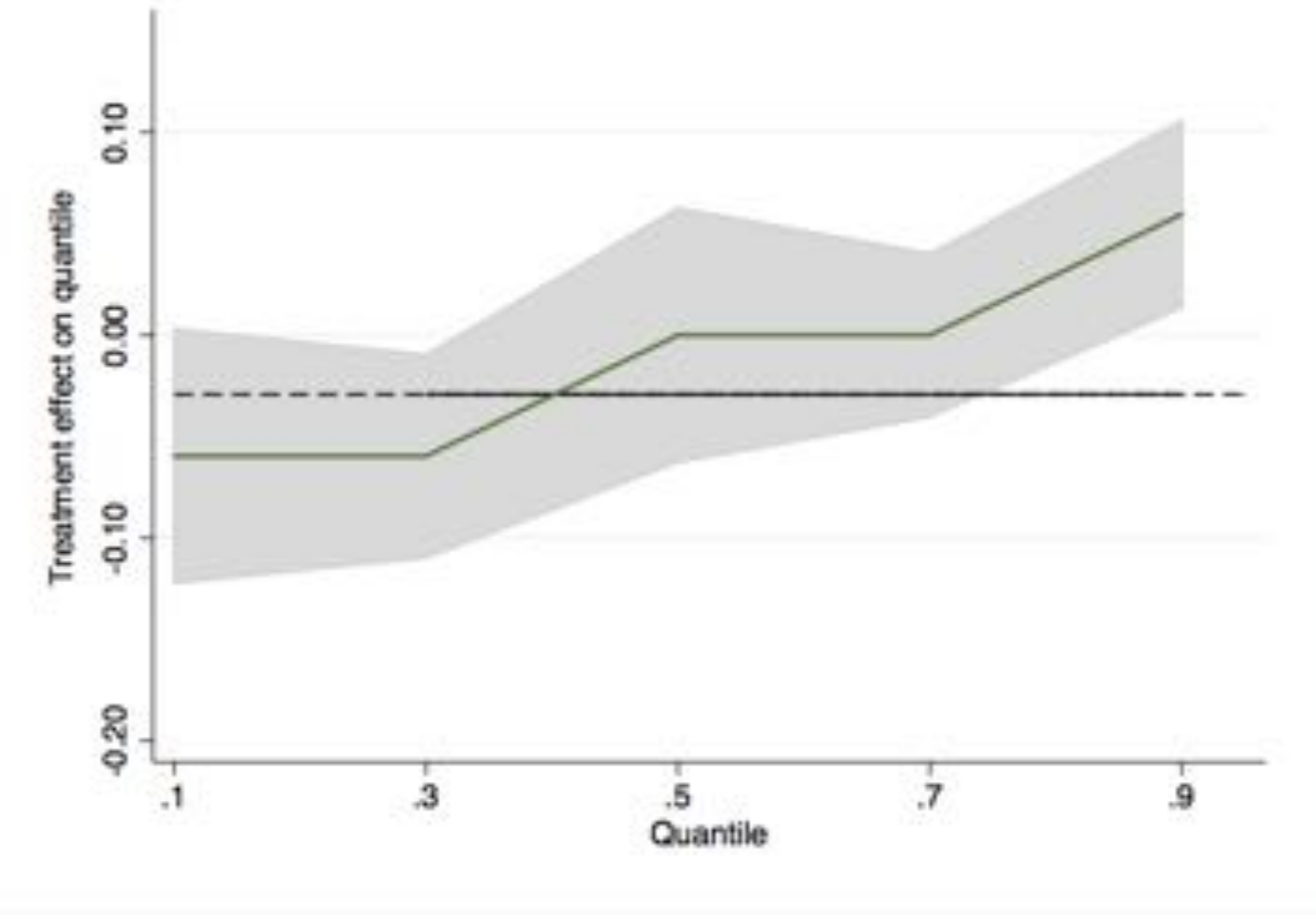

Figure 8: Treatment effect on composite student test scores by quantiles

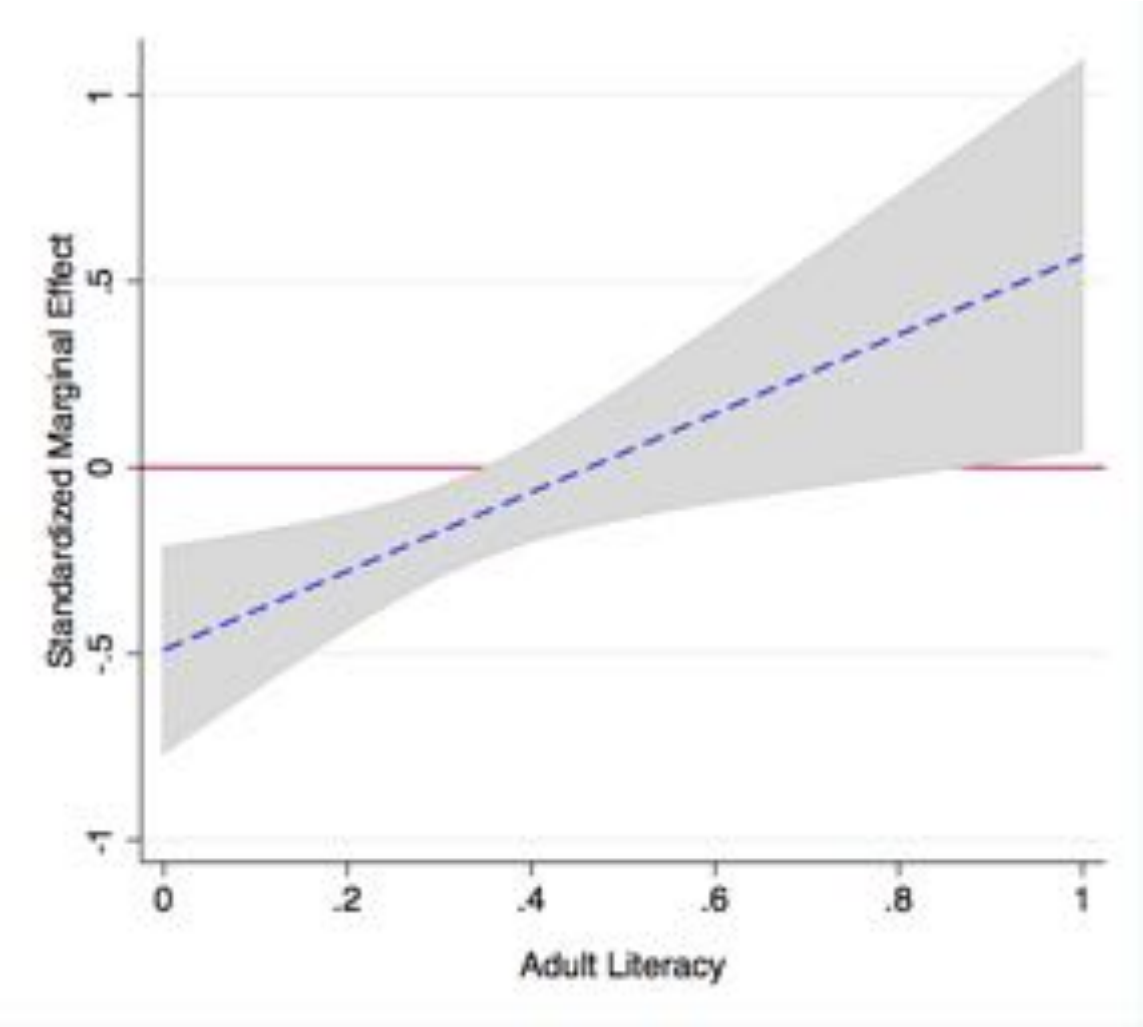

Figure 9: Level of baseline adult literacy and effectiveness of the WSD on composite student test scores

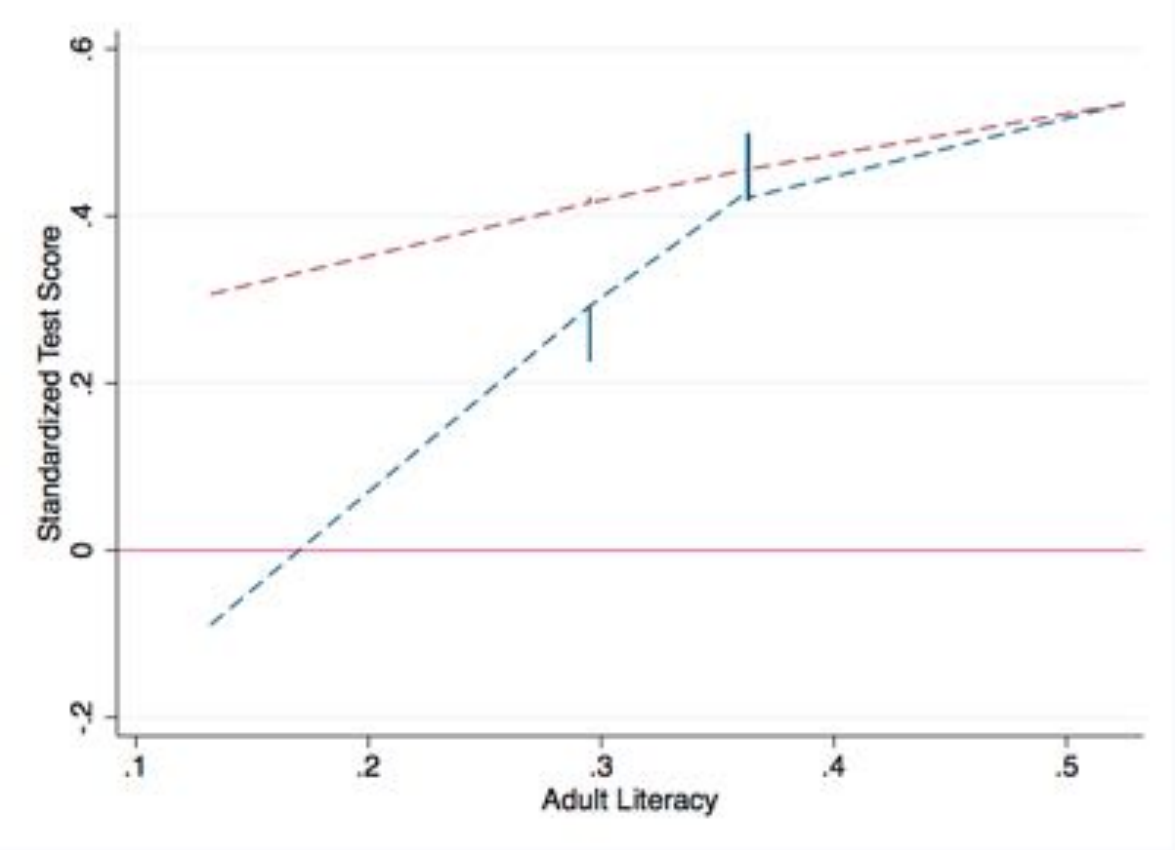

Figure 10: Level of baseline adult literacy and effectiveness of the WSD on composite student test scores: Non-parametric estimate

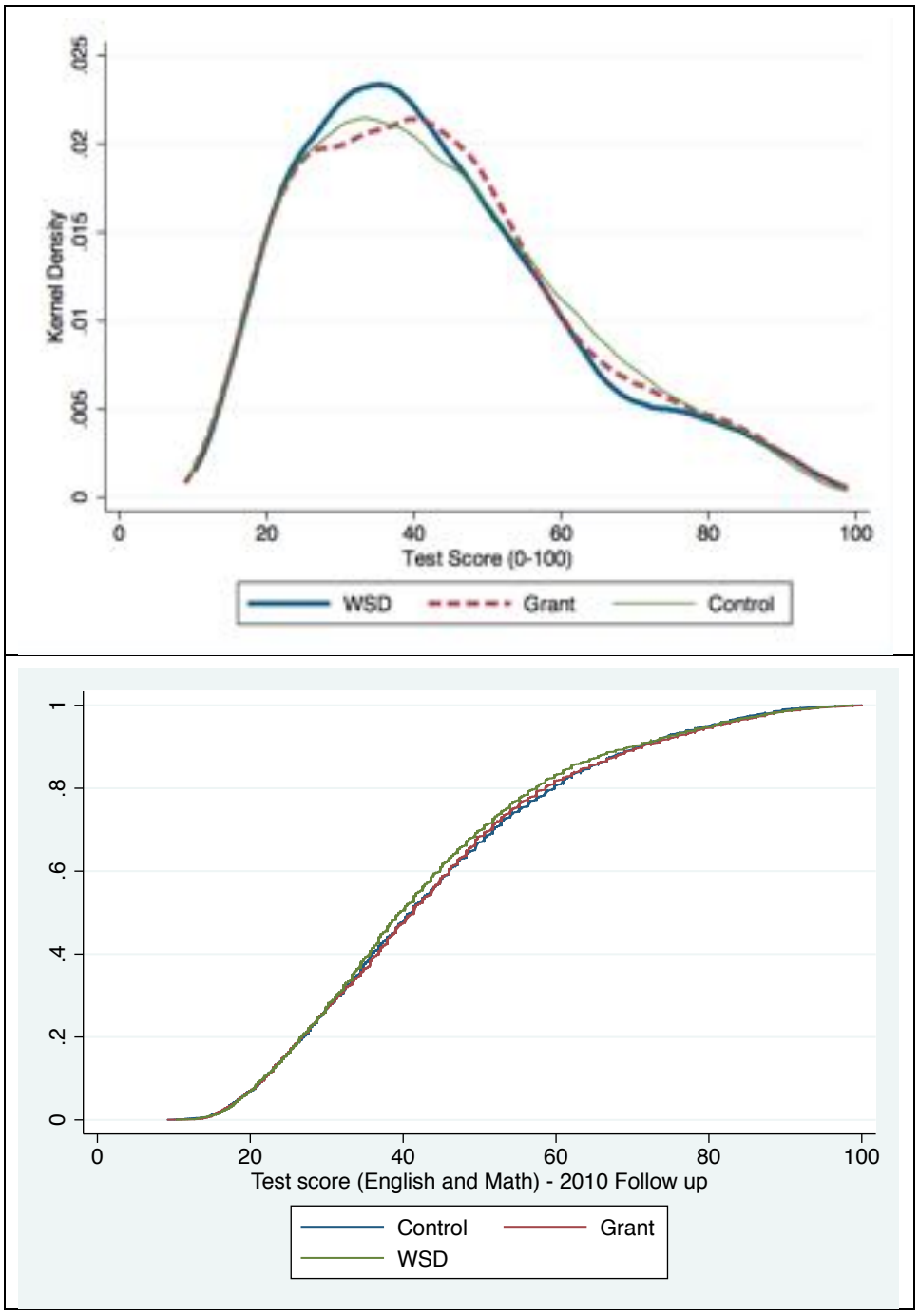

Figure 11: Distribution of composite test scores (2010)

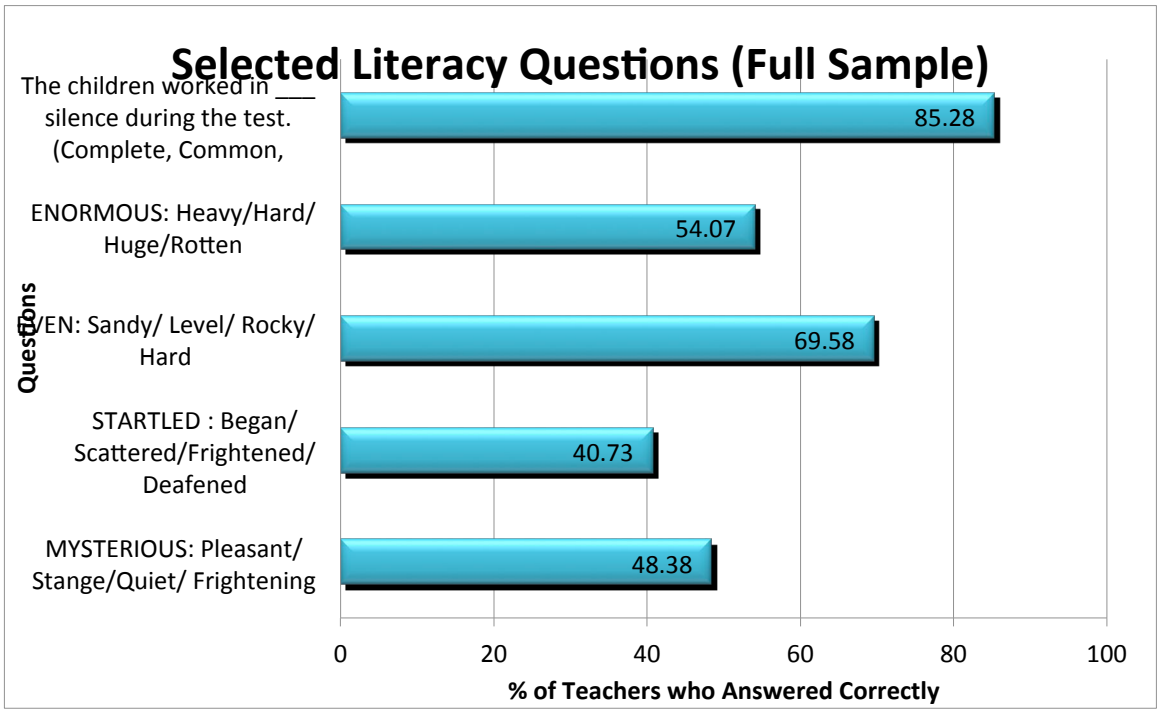

Figure 12: Teacher Content knowledge on selected English questions

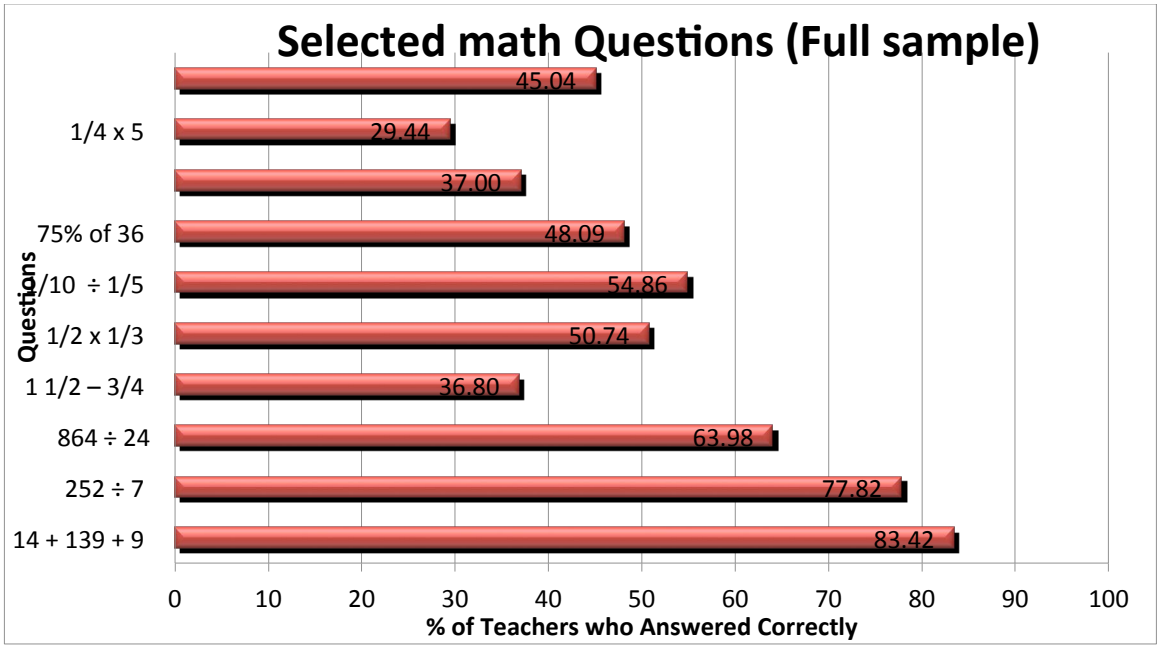

Figure 13: Teacher content knowledge on selected math questions

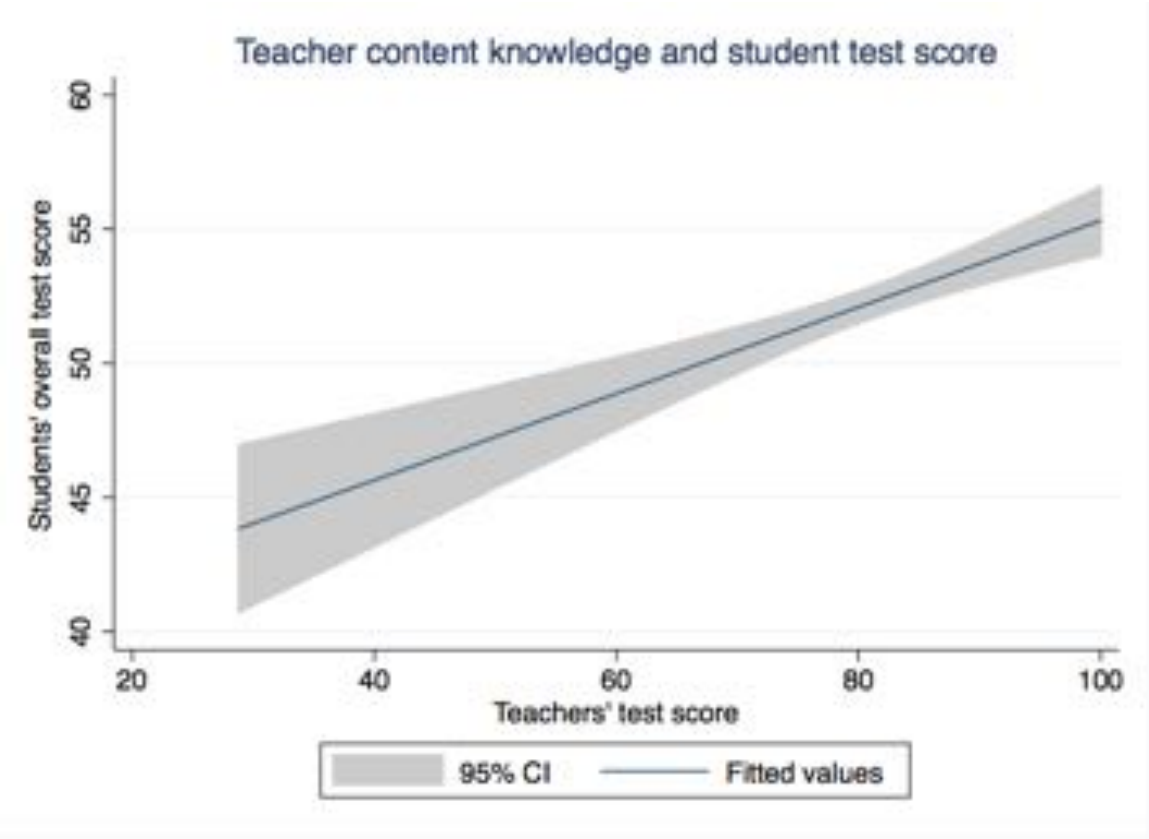

Figure 14: Correlation between student and teacher test scores