Inputs, Incentives, and Complementarities in Primary Education: Experimental Evidence from Tanzania

Isaac Mbiti, Karthik Muralidharan, and Youdi Schipper[1]

Preliminary and Incomplete- Do Not Cite

Abstract

Recent nationwide assessments have documented the low levels of learning in Tanzanian schools. These low levels of learning are driven in part by limited accountability in the education system, which is reflected in the frequent absence of teachers from schools. This is further compounded by the resource constraints that schools face. In this study we conduct a randomized experiment to examine the efficacy of increasing resources to schools relative to increasing teacher incentives. Specifically, we compare the student learning outcomes between four different interventions: one in which we provide schools with extra resources through capitation (or per pupil) grants, one in which we provide teachers with a bonus based on the performance of their students in an externally administered exam, one in which schools received both programs, and the control group which received no support. Overall, we find limited evidence that solely providing resources improves learning outcomes, while we do find some evidence that incentives improve learning outcomes especially when coupled with extra resources

[1] Preliminary and incomplete. Do not cite. Usual disclaimer applies. Author Affiliations: University of Virginia, University of California- San Diego, and Twaweza.

Introduction

Over the past decade, developing countries have made significant investments aimed at increasing access at primary education. By employing a variety of programs such as conditional cash transfers, school feeding programs, school based health programs, and fee reductions the net enrollment rates in developing countries have increased dramatically over the past two decades. East African countries have seen sharp increases in their primary enrollment rates, in part due to their free primary education programs (Lucas and Mbiti, 2012 in Kenya, Grogan, 2010 in Uganda, and Valente, 2014 in Tanzania). In Tanzania, the setting of this study, the net enrollment rate in primary school rose from 53% in 2000 to close to 90% in 2012 (World Bank, 2014). Yet despite Tanzania's progress toward achieving near universal primary school access, there is a growing concern that the quality of education may have been compromised. Recent independent nationwide schooling assessments highlighted the low levels of learning in Tanzanian schools, where less than one third of third graders could demonstrate competency in second grade numeracy or literacy (Uwezo, 2013). The typical policy response has emphasized the need to alleviate resource constraints in schools often through capitation (or per-pupil block) grants to schools. However, a large body of evidence suggests that increased monetary resources do not improve student learning outcomes (see McEwan, 2014; Murnane and Ganimian, 2014, Kremer, Brannen and Glennerster, 2013). These systematic reviews further suggest that that most effective education interventions are ones that change students' classroom experience.

Teacher incentives (or performance pay) could change the classroom experience, especially in low accountability settings such as Tanzania (Murnane and Ganimian, 2014). These low levels of accountability are reflected in the high levels of teacher absenteeism from classrooms, with public teachers only spending 2 hours/day in Class (World Bank Service Delivery indicators, 2010). While papers such as Muralidharan and Sundararaman (2011) show that teacher incentives can improve learning outcomes, other students, such as Fryer (2012) or Kremer et al (2010) find limited evidence of the efficacy of such programs. The differences in the effectiveness of teacher incentives is partly driven by the markedly different incentive designs that employed in these studies. Muralidharan and Sundararaman (2011) evaluated a teacher performance pay program that rewarded teachers on the basis of student value added, whereas the programs in Fryer (2012) and Kremer et al (2010) rewarded teachers if their students demonstrated a certain pre-determined competency level. A drawback of rewarding teachers on the basis their students' achieving a competency level (e.g. passing a test) is that

teachers may not direct as much effort towards students who are far away from the threshold. This implies that the best students and the worst students may not be well served by such incentive designs. On the other hand, these types of incentive programs are commonly implemented as they are simpler to understand, implement and scale-up, whereas programs that reward teachers for value-added are much harder to understand and scale-up as they require a complex student test-score tracking system[2].

Given the limited capacities of countries such as Tanzania to scale up more complex (and "more optimal") incentive programs, simple incentive programs that use a threshold design could potentially be a cost-effective approach to raise the productivity of the education sector. Such a program could encourage teachers to increase their efforts, which could result in decreased absenteeism and increased in time spent teaching. Moreover, such a program could encourage teachers, head-teachers and other administrators to use their available resources more effectively. For example research by Evans and Sabarwal (2014) showed that textbooks are generally kept locked up in store rooms, rather than in active use by students in class. An incentive program could potentially encourage teachers to more effectively utilize textbooks by incorporating them in their daily lessons and even allowing the children to take them home. However, there is very limited evidence on the complementarities of inputs (or resources) and incentives. Given that programs (or interventions) occur in the context of an education system, a better understanding of the complementarities between proposed programs (or interventions) is extremely important for policymakers.

We use randomized control trial to compare the effectiveness of alleviating (monetary) resource constraints to the effectiveness of introducing teacher performance pay. The RCT is designed to further examine the complementarities between inputs and incentives. We sample a nationally representative set of 350 schools across 10 districts in Tanzania. We the randomly allocate our set of 350 schools to four groups: Group 1 schools receive Capitation Grants (per pupil grants), Group 2 schools receive a simple performance pay program, Group 3 schools receive both grants and incentives, and Group 4 schools are our control schools. Although primary schools in Tanzania teach several subjects from grade 1 to grade 7, our study only focuses on student performance in Math, English and Kiswahili in grades 1, 2 and 3. Consistent with previous studies, we find that merely increasing resources in schools does not lead to improved learning outcomes, even though the capitation grant program nearly doubled (non-teacher) spending per child in group 1 and group 3 treatment schools. We also find that the incentive

[2] No Child Left Behind is an example of a program that introduces threshold effects due to its focus on students achieving a certain qualification mark.

program does not yield statistically significant gains in learning, although the coefficients are positive. We do however find that the combination of incentives and resources led to a positive and significant increase in learning. Relative to the control group, student test scores in combination schools increased by 0.20 SD. Additional statistical tests show that the treatment effect in combination schools is larger than the sum of the estimated treatment effects for capitation grant schools and incentive schools. We argue that this suggests that there are complementarities (or synergies) between incentives and resources. Our findings illustrate the importance of designing RCTs to have sufficient power to detect complementarities. They also highlight the danger of ignoring complementarities in RCTs with multiple arms, especially RCTs with cross- cutting designs, as this may yield misleading findings.

Background and Context

Tanzanian primary schools serve students from first grade (or standard one) through seventh grade (or standard seven). Students take two national examinations, one at fourth grade and another at seventh grade. The seventh grade exam is the primary school exit examination and determines if children can progress to secondary school. The fourth grade exam is "low-stakes" and is generally used to for administrative and monitoring purposes. In 2001, Tanzania abolished primary school fees in public primary schools. The previously collected fees were replaced with a capitation grant (or per-pupil block grant to schools) that would support the operation of schools. The official policy stipulated that schools should receive approximately US$10 per enrolled student. The policy also provided spending guidelines where 40% of these grants could be spent on textbooks, teaching guides and other reading materials, administrative expenses, 20% could be spent on chalk, exercise books, pens and pencils, 20% on minor construction and repair, 10% on examinations and test paper printing and 10% on administration (Twaweza, 2010). As teachers were assigned to schools and paid directly by the government, schools could not use capitation grant funds to pay teachers or hire teachers.

While the official policy stipulates that schools should receive TZS 10,000 (around US$6.25) per student, the Tanzanian government has generally failed to provide that level of support to schools. From 2006-2010, schools were only allocated about TZS 6,000 (US$3.75) per student by government (Twaweza, 2010). School finances were further constrained by significant leakages of the budgeted capitation funds. A recent World Bank study estimated that about 37% of budgeted capitation grant funds did not reach the school (World Bank SDI, 2012). This leakage is in part driven by the disbursement system, where funds are first transferred to regional and local authorities who are then responsible for transferring the funds to schools.

As a result of the limited finances available to schools, the educational inputs and infrastructure at schools is often inadequate. The World Bank Service Delivery Indicators show that only 3% of schools have sufficient infrastructure (potable water, sanitation, and electricity) and 5 children (in grades 1, 2, and 3) shared a math textbook, while 2.5 children shared a reading book (World Bank EDStats, 2012). The increase in enrollment spurred by the introduction of the free primary school program has resulted in large class sizes and large pupil to teacher ratios (Valente, 2015). Class sizes in primary schools average 74 students, with much larger classes in lower grades. There are also almost 50 students per teacher in primary schools (World Bank SDI, 2012).

In addition to large class sizes and limited resources, there is also limited accountability in Tanzanian public primary schools. Teacher absence rates are high. Almost one in four teachers is absent from school on a given day (World Bank SDI, 2012). Teacher effort also seems low even among those that are present in school. Over 50% of teachers who were present in school were absent from the classroom (World Bank SDI, 2012). As a result of the low levels of teacher effort, children only received about 2 hours of instruction per day at schools. These low levels of effort could reflect low levels of teacher motivation. A recent UNICEF Tanzania report supported the introduction of policies that rewarded teachers for quality teaching (UNICEF, 2013). Through the Big Results Now initiative, the government has prioritized programs that increase teacher motivation and is piloting programs that reward teachers and schools that can deliver better learning outcomes. This policy shift is notable as it marks the first time that the Tanzanian education system is focusing on learning outcomes rather than educational inputs.

Tanzanian learning outcomes are abysmal. Nationwide assessments from the Uwezo surveys show that less than one third of grade 3 students could read at a grade 2 level in Kiswahili (the national language and language of instruction) or successfully demonstrate grade 2 numerical skills. Performance in English was especially weak, with less than 12% of grade 3 students able to read at a grade 2 level in English.

Design

We conduct our RCT in nationally representative sample of 350 schools across 10 districts in Tanzania (see Figure 1). In each of our 10 study districts, we randomly assign 7 schools to receive the capitation grant intervention (CG), 7 schools are assigned to receive teacher incentives ("Cash on Delivery" or COD), 7 schools receive both grants and incentives (Combo) and 14 schools are assigned to the control group. The double sized control group increases our power to detect complementarities between incentives and resources. Thus, our full sample consists of 70 schools in the capitation grant arm, 70 in

the performance pay program, 70 in the combination arm and 140 in the control group. The intervention was managed by Twaweza, a Tanzanian NGO that focuses on citizen agency and public service delivery. Intervention schools were informed that the program would last for two years. Schools were informed about the program through community meetings. Program materials such as flyers and cartoon booklets were distributed to teachers and to students who were instructed to share the materials with their parents. Twaweza also worked closely with both central and regional government officials to ensure that there would be minimal interference with the program and the research efforts.

Capitation Grant (CG) Arm

Schools assigned to receive the capitation grant were provided with the grants that mirrored the official Tanzanian capitation grant policy. Thus each school received TZS 10,000 (approx. US$6.25) per student from Twaweza. Schools receiving these funds had to spend and account for the funds as outlined in the official policy (described above). Thus they could spend 40% of the funds on textbooks but not spend any of the funds on teachers. As discussed above, capitation grant leakages were a major challenge faced by schools. In addition the timing of government grant disbursements was very unpredictable making financial planning by schools extremely difficult (Twaweza, 2010). In order to minimize leakage and enable better financial planning, Twaweza grants were transferred directly into school bank accounts in two tranches, the first at the beginning of the second term (around April) and the second at the beginning of the third term (around August/September).

Following the regular capitation grant policy, schools were also required to share revenue and expenditure information with the community and display summary financials in a public area in the school (usually on a notice board in the school). On aggregate, approximately US$700,000 was disbursed to these schools each year in 2013 and 2014 (this includes schools in the CG and Combination arm). The size of the grants distributed to schools is approximately three times the pre-treatment per student expenditure (excluding teacher salaries). Thus this reflects a significant increase in the financial resources available to schools. The successful implementation of this intervention prompted the government to review it capitation grant policy. The government now plans to transfer capitation grant funds directly to schools (The Guardian (TZ) Newspaper, 2014).

Cash on Delivery (COD) Arm

The teacher performance pay program provided monetary bonuses to teachers contingent on the performance of their students. Given Twaweza's emphasis on early grade learning, the program was limited to teachers in grades 1, 2, and 3 and focused on numeracy and literacy in English and Kiswahili.

Individual eligible teachers would earn a TZS 5,000 (US$3.125) bonus for every student that passed a simple externally administered grade appropriate assessment (based on the curriculum) in Math, English, and Kiswahili. If a teacher taught all three subjects in grade one, for example, they would receive TZS 15,000 (US$9.375) for every student that passed all three assessments. In addition to Math, English and Kiswahili, students in grades one, two, and three studied several subjects including technology, vocational studies (such as agriculture and home science), and science. In some schools students had the same teacher for all subjects, but in other cases students had different teachers for each subject. We also included head teachers in the incentive program. Head teachers would earn TZS 1,000 (US$ 0.625) per student that passed an assessment.

The program was announced to teachers in March of 2013 and a follow up visit in July 2013 reinforced the details of the program. The assessments were scheduled for the end of the school year and in many cases served as the end of year exams for schools. Understanding of the program was high as over 90% of teachers in the program could correctly calculate the bonus level in a hypothetical scenario. The simplicity of the incentive program enabled teachers to understand the program. The simple "threshold design" utilized here was also simpler to implement and is arguably more scalable in resource (both human and financial) constrained settings such as Tanzania. However, threshold designs are not optimal incentive designs as they can encourage teachers to focus their attention on students close to the passing threshold. Thus students who are far below the passing threshold and those who are far above the threshold would not be well served by such an incentive design. Despite the limitation of this design it is important to note that it is routinely used in programs such as No Child Left Behind (Neal and Schazenbach, 2010).

In order to ensure the fidelity of the implementation, we created several versions of the high stakes end of year test. We also took student photos at enrollment to prevent identity fraud. We also conducted the end of year testing in a sample of control schools and also conducted a low stakes audit test (or research test) on a sample of 30 students (10 students in each grade) in all 350 schools in our sample. Teachers (and head teachers) earned about $150,000 in bonuses each year (this includes teachers in the COD only and Combo arm). These bonuses were paid directly into teacher bank accounts or through mobile money transfers.

The table below summarizes the main features of the interventions.

	Capitation Grant (CG)	Performance Pay (COD)
Amount	10,000 TZSHS per student enrolled in school	5,000 TZSHS awarded to teacher conditional on student passing externally administered exam. 1,000 TZSHS also awarded to Head teacher if student passes.
Grades eligible	Awarded based on enrollment in entire school	Only teachers in Grades 1,2,3 are eligible.
Subjects eligible	N/A	Only teacher teaching Kiswahili, English and Math in relevant grades are eligible
Restrictions on use of funds?	Funds are given to school. Can be used on administration, books, teaching materials, repairs or small construction (e.g. build toilets), casual laborers (e.g. security guard, cooks). Cannot be used to pay teachers. School management committees responsible for managing these funds.	Funds are given directly to teachers. No restrictions on how they use it.

Combination Arm

Schools assigned to the combination arm received both capitation grants and teacher incentives. As discussed earlier we increased the size of the control group in order to increase our ability to detect complementarities between incentives and resources. Figure 2 outlines the timing of the research and intervention activities including the data collection and the information/dissemination meetings and end line testing.

Data

From each school, we sample 10 students from each grade focal grade (grade 1, 2 and 3) to create a student panel of 10,000 students who we follow over the course of this study. This provides us with a panel of student test scores. From this set of students, we randomly sample 3500 students to conduct household surveys. These surveys collect information on educational expenditures, general household characteristics and non-financial educational inputs at the household (such as helping with homework). We survey all teachers (about 1500) who teach focal grade (grade 1, 2, 3) and focal subjects (Math, English and Swahili). We measure teacher effort, teacher time use and teaching strategies. We also obtain teacher opinions on incentive programs. We further conduct 350 head teacher/ school level surveys. We collect information about the school facilities, the teaching roster, input availability and expenditures. Unlike most school inputs, textbooks are easily assignable to grades and subjects. We therefore collect information on textbook purchases at the grade subject level. In addition to this survey data, we also utilize administrative data from the program, including test scores from the external test that is used to assess students for the COD program.

The baseline was conducted in early 2013, followed by a midline in August 2013 and an end line in October 2013. The pay for performance testing occurred in November 2013. A similar calendar was followed in 2014 (see Figure 2). Prior to analysis the data, we filled a pre-analysis plan on the AEA registry.

Results

Tables 1-3 shows that the observable characteristics are generally balanced across our treatment arms. We normalize test scores relative to the mean in the control group in each grade- subject. Table 2 shows that schools had limited infrastructure and had large enrollments with on average 50 students per teacher. Less than 10% of schools tracked students and 40% of schools had multiple shifts, where half the grade would attend school in the morning and the other half would attend from the late morning. Schools (and households) were mostly rural. Households spent about US$8 on the focal child's education and the majority of these households had floors made of earth/mud (rather than cement or more durable materials). Table 3 shows that about 60% of teachers in our sample were female and had about 15 years of experience overall and 7 years of experience at their current schools.

Capitation Grant

Table 4 shows how schools receiving the grants spent the funds across broad categories. Overall there are no statistically significant differences in spending behavior between combo schools and regular capitation grant schools. In the first year schools spent about 80% of the grant and saved the remainder. Schools were more restrained in spending the grant in year two. Given the uncertainties of government funding (both in terms of timing and amount), this "precautionary saving" behavior by school is quite reasonable. Table 5 examines total school expenditures across all sources. In both years spending at Combo schools and CG schools was approximately two times that of control group schools. In general spending patterns between Combo and CG schools were similar. However, in Year 1 we see suggestive evidence of complementarities where Combo schools invest more in teaching aids such as wall charts, than the combined spending by incentive schools and CG schools (Table 1, Panel A, Column 4). This pattern could reflect lobbying efforts by teachers in grade 1, 2 and, 3 to obtain resources that could improve their likelihood of earning bonuses.

A potential drawback of capitation grant programs that are provided by NGOs is that households, governments and other stakeholders may cut back their support, thereby undermining the program (Das et al, 2013). Table 6 shows that there were moderate cutbacks in capitation grant schools in the second year but not the first year. Parents cut back expenditures by about TZS 1,000 in both CG and Combo schools but this was not statistically significant. Table 7 shows that households cut back slightly on textbooks spending in year 2. Overall, the minor reductions in resource support from other support did not fully offset the Twaweza grant, thus the capitation grants did lead to overall increases in educational support and spending.

Test Scores

Table 8 below shows the main results of our study. Focusing mainly on the second year of results, we see that capitation grants (CG) do not raise test-scores. We also see that the performance pay program (COD) is not effective in raising test-scores in isolation. However, we see that the combination of both programs significantly raises learning outcomes, especially in Kiswahili and Math. For robustness and to mitigate concerns of multiple testing, we create summary index of focal subjects. Pooling these subjects, we see that the combination arm increases test scores by almost 0.2 SD relative to the control group. We formally test whether the treatment effect of the combination arm is greater than the sum individual CG and COD arms and report the P-Value of that test. Our tests show that there is significant evidence of the complementarities of resources of incentives especially in Swahili and Math.

We examine if there are positive (or negative) spillovers into non-incentivized subject (science) in Table 9. We do not find any evidence that the gains reported in Table 8 above came at the expense of other subjects. Overall the results suggest there are likely positive spillovers as students may be able to better understand other subjects as a result of improvements in literacy and numeracy. We also examine if our interventions affected performance in the Grade 7 national exit examination. One potential concern is that schools may invest most of their funds to support Grade 7 since school reputations are based on performance in this exam. However we do not see any evidence that any of our interventions significantly affect performance in the national exit exam.

Teacher effort

We use self-reported data from teacher to examine differences in teacher effort. Teachers could offer remedial help to students who are falling behind, conduct more testing in class to prepare students and also offer tutoring after school (for a fee). In the first year we see that teachers in the incentive program gave more tests and were more likely to offer tutoring. However we do not see similar behavior among teachers in the combination arm. In the second year we see that teachers in combination schools held more tutoring sessions. Table 11 examines time use by teachers. In the first year teachers in COD and Combo schools devoted more time to extra classes, however in year two we do not see any systematic changes in time use patterns. We argue that these results suggest that teacher behavior did change in response to the incentives, but our self-reported measures are able to yield consistent clear patterns of behavior change.

Heterogeneity

We explore heterogeneous treatment effects by gender, age and baseline test scores in Table 12. Overall there is limited evidence of heterogeneous effects by these characteristics. Although we note that girls performance in English improves less in Combo and COD schools.

As discussed earlier, threshold incentive designs can lead teachers to focus on students near the passing level, perhaps at the expense of students who are in the tails of the distribution. We explore the potential for these effects in Tables 3-8. Using non-parametric analysis we find evidence that the effects of the program (in Combo schools) were concentrated among students in the “middle” of the distribution, especially for Math and Swahili. This is consistent with the program introducing threshold effects, where teachers only focus on students near the qualification threshold. Given the low levels of English language skills, the treatment effects for students (in the Combo arm) in English were

concentrated in the right tail. This is again consistent with the notion that teachers would only focus on students who were at the margin of passing.

Discussion

Our results highlight the importance of accounting for complementarities in study design. Cross-cutting research designs are commonly used to achieve cost savings. In a 2003 article Kremer states "Since data collection is the most costly element of these evaluations, cross-cutting the sample reduces costs dramatically… This tactic can be problematic, however, if there are significant interactions between programs" (Kremer, 2003). Table 13 highlights the danger of ignoring the interaction effects between programs. Column 1 of Table 13 assumes the interaction effects are zero and presents estimates of the treatment effect of the capitation grant and the incentive program. Focusing on Panel C (our pooled estimates) we see that our inference would be misleading as we would conclude that both programs yielded statistically significant increases in learning (relative to the control group).

Conclusion

In this paper we report findings from a large education RCT aimed at improving learning in early grades. Consistent with other findings, we show that merely increasing school resources do little to improve learning outcomes. We also find that a simple incentive program yield insignificant but positive impacts on learning. We find that test scores in schools that received both programs were significantly higher. Moreover, we find strong evidence of complementarities between inputs/resources and incentives. We further find that the increases in learning (in the combo schools) were concentrated among students near the passing threshold. This is further evidence of the importance of incentive design in promoting student learning. Finally, we highlight the potential danger in inference in ignoring complementarities between programs. Ignoring complementarities in cross-cutting experimental designs may yield biased estimates and lead to the scale up and adoption of ineffective programs. We thus argue that researchers should adequately design studies to account for complementarities.

1 Figures

Figure 1: Districts in Tanzania from which schools are selected

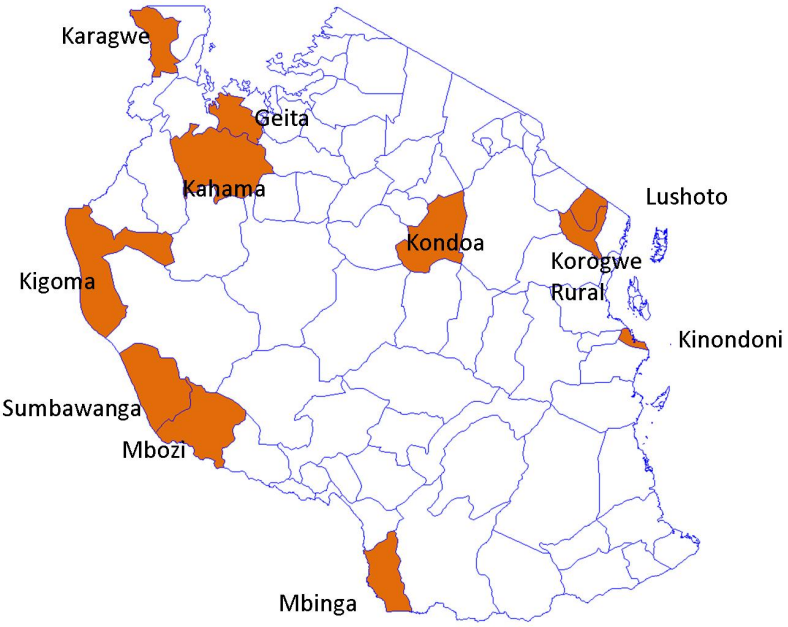


Figure 2: Timeline

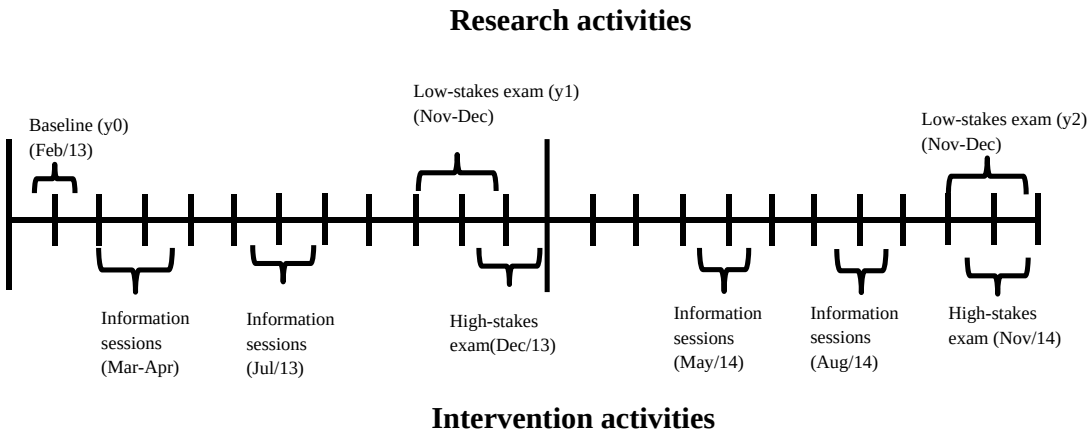
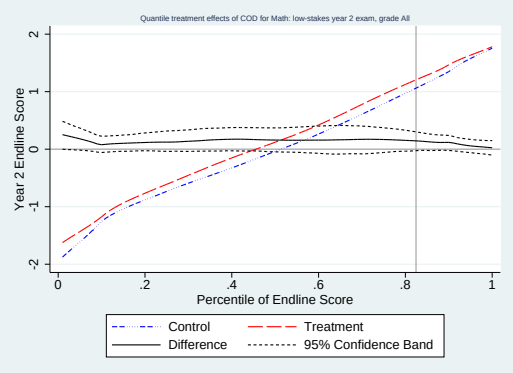


Figure 3: Effect of Cash on Delivery (left) and Combo (right) on Math test scores by endline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.



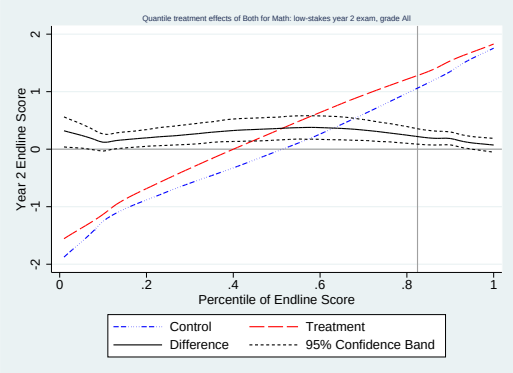
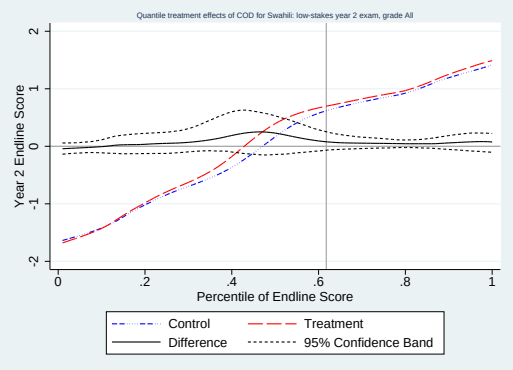


Figure 4: Effect of Cash on Delivery (left) and Combo (right) on Swahili test scores by endline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.



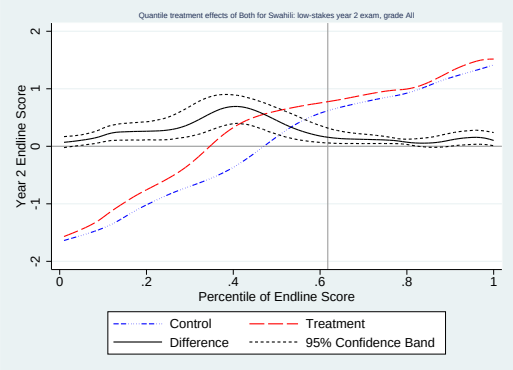


Figure 5: Effect of Cash on Delivery (left) and Combo (right) on English test scores by endline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.

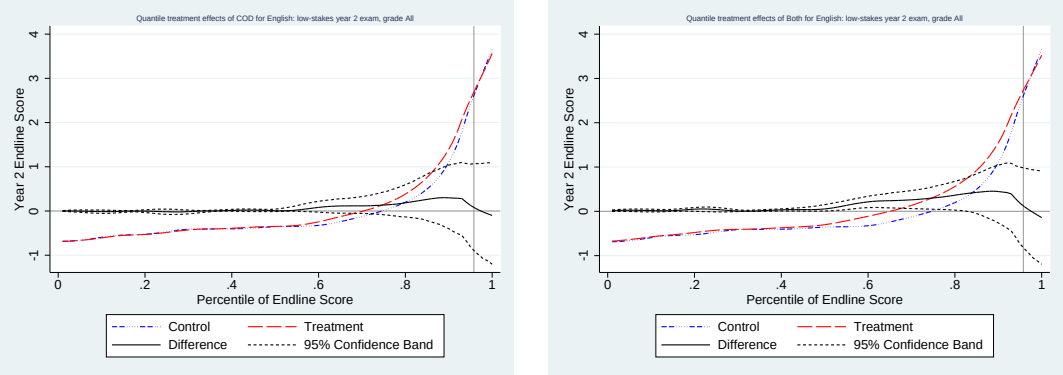


Figure 6: Effect of Cash on Delivery (left) and Combo (right) on Math test scores by baseline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.

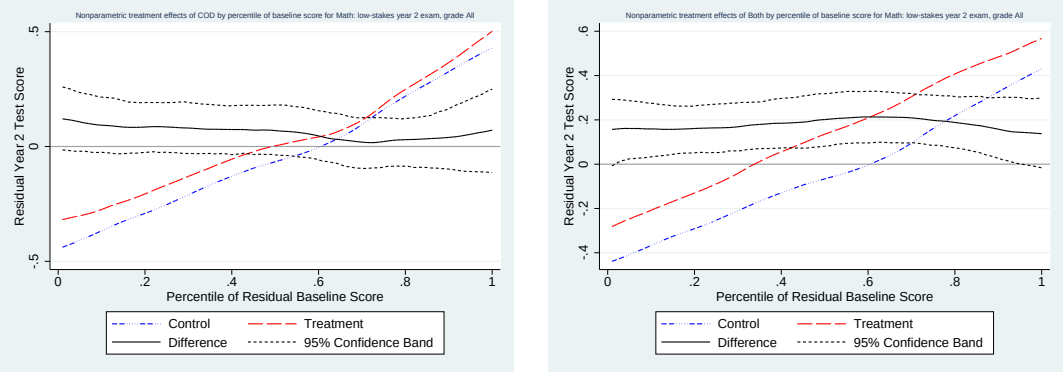
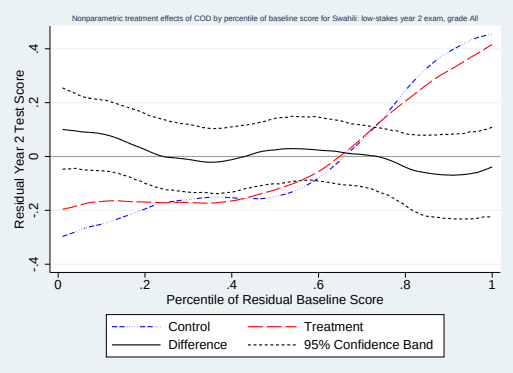


Figure 7: Effect of Cash on Delivery (left) and Combo (right) on Swahili test scores by baseline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.



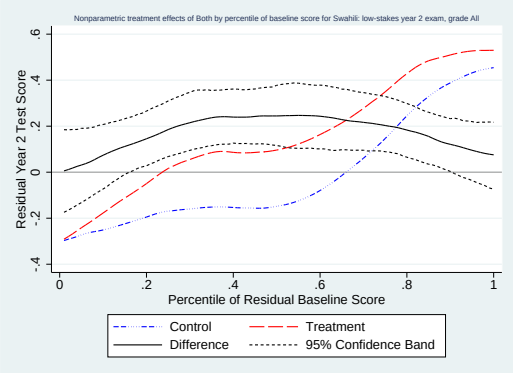
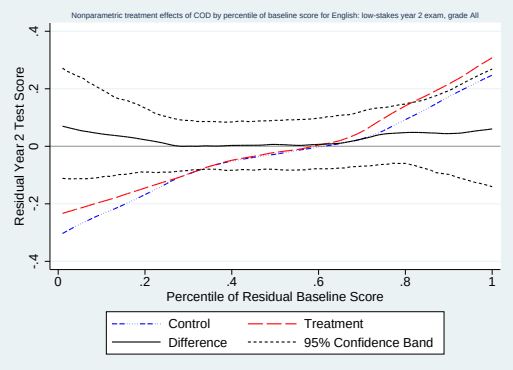
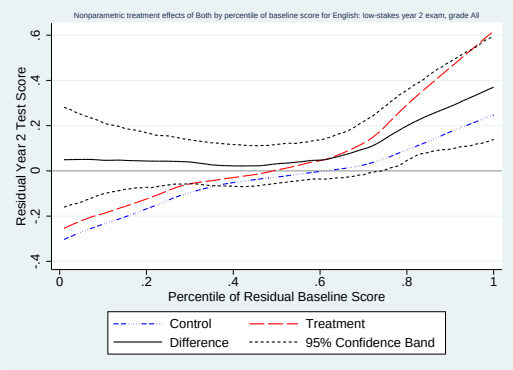


Figure 8: Effect of Cash on Delivery (left) and Combo (right) on English test scores by baseline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.





2 Tables

Table 1: Baseline descriptive/balance: students

	Combo	CG	COD	Control	p-value (all equal)
Seen Uwezo Test	0.021	0.016	0.017	0.016	0.68
	(0.0038)	(0.0035)	(0.0039)	(0.0025)	
Went to Preschool	0.81	0.75	0.78	0.77	0.42
	(0.023)	(0.024)	(0.023)	(0.018)	
Male	0.50	0.49	0.50	0.50	0.99
	(0.0095)	(0.010)	(0.0085)	(0.0075)	
Age	8.94	8.96	8.94	8.96	0.96
	(0.050)	(0.052)	(0.047)	(0.039)	
Swahili test score	0.053	-0.020	0.065	0.00014	0.78
	(0.067)	(0.067)	(0.082)	(0.048)	
Math test score	0.069	0.0076	0.059	0.00016	0.77
	(0.061)	(0.063)	(0.072)	(0.045)	
English test score	-0.018	-0.013	0.020	0.00010	0.96
	(0.047)	(0.049)	(0.063)	(0.037)	

Standard errors, clustered at the school level, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Baseline descriptives/balance: schools

	Combo	CG	COD	Control	p-value (all equal)
Panel A: Schools					
Infrastructure Index (0-6)	2.24	2.33	2.34	2.49	0.20
	(0.094)	(0.095)	(0.097)	(0.073)	
Computers	0.014	0.043	0.043	0.064	0.23
	(0.014)	(0.024)	(0.024)	(0.021)	
Electricity	0.13	0.14	0.13	0.14	0.99
	(0.040)	(0.042)	(0.040)	(0.029)	
Single shift	0.60	0.59	0.64	0.63	0.89
	(0.059)	(0.059)	(0.058)	(0.041)	
Teachers/Students	0.021	0.020	0.021	0.021	0.96
	(0.00096)	(0.0011)	(0.0010)	(0.00077)	
Track students	0.071	0.10	0.071	0.093	0.88
	(0.031)	(0.036)	(0.031)	(0.025)	
Urban	0.16	0.13	0.17	0.15	0.91
	(0.044)	(0.040)	(0.045)	(0.030)	
Enrolled students	739.1	747.6	748.5	712.4	0.89
	(48.4)	(51.9)	(51.7)	(30.4)	
Panel B: Households					
Household size	6.247	6.332	6.436	6.352	0.805
	(0.134)	(0.141)	(0.137)	(0.0940)	
Asset Index (0-8)	1.107	1.078	1.111	1.104	0.956
	(0.0454)	(0.0468)	(0.0466)	(0.0318)	
Expenditure in education (2013)	12909.4	12180.1	13126.8	14948.1	0.494
	(1468.1)	(1422.1)	(1587.0)	(1254.0)	
Walls made out of earth/mud	0.459	0.404	0.416	0.438	0.782
	(0.0405)	(0.0414)	(0.0418)	(0.0294)	
Floor made out of earth/mud	0.657	0.659	0.660	0.668	0.994
	(0.0365)	(0.0367)	(0.0367)	(0.0249)	
Roof made out of a durable material	0.804	0.805	0.793	0.801	0.989
	(0.0293)	(0.0225)	(0.0289)	(0.0186)	

Standard errors, clustered at the school level, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Baseline descriptives/balance: teachers

	Combo	CG	COD	Control	p-value (all equal)
teacher gender	1.64	1.67	1.67	1.65	0.89
	(0.039)	(0.041)	(0.035)	(0.026)	
year born	1974.7	1975.2	1974.9	1975.1	0.98
	(0.76)	(0.79)	(0.65)	(0.48)	
year start teaching	1998.8	1998.8	1998.7	1999.0	0.99
	(0.79)	(0.83)	(0.68)	(0.49)	
year start teaching at this school	2006.4	2005.9	2006.4	2006.3	0.84
	(0.46)	(0.45)	(0.45)	(0.31)	
Taught in a private school	0.032	0.028	0.026	0.033	0.88
	(0.0078)	(0.0065)	(0.0067)	(0.0060)	

Standard errors, clustered at the school level, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: How are schools spending the money?

	Combo	CG	Diff
Year 1			
Total	8118.5	7962.2	156.3
	(296.3)	(333.9)	
$ Admin./Student	1995.2	1773.1	222.2
	(139.0)	(148.3)	
$ Student/Student	450.5	622.5	-172.0
	(82.64)	(94.69)	
$ Teaching Aid/Student	5803.9	5620.1	183.7
	(205.9)	(285.0)	
$ Teacher/Student	2.742	0	2.742
	(1.968)	(0)	
$ Construction/Student	98.13	60.35	37.78
	(51.42)	(36.58)	
Year 2			
Total	5453.8	5981.9	-528.1
	(359.3)	(353.6)	
$ Admin./Student	2023.3	2111.0	-87.72
	(167.9)	(200.7)	
$ Student/Student	409.0	449.8	-40.73
	(65.03)	(81.17)	
$ Teaching Aid/Student	3110.0	3434.5	-324.5
	(259.9)	(265.1)	
$ Teacher/Student	0	3.316	-3.316
	(0)	(3.316)	
$ Construction/Student	67.31	68.76	-1.448
	(39.29)	(60.29)	
$ Savings/Student	4102.7	3620.9	481.8
	(328.5)	(365.3)	

Mean expenditure per student. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Total expenditure

	$ Total.	$ Admin.	$ Student	$ Teaching Aid	$ Teacher	$ Construction
Panel A: Year 1						
CG	5569.1***	1546.9***	826.1***	3039.6***	-43.9	200.5
	(812.0)	(405.8)	(125.8)	(399.3)	(81.2)	(414.7)
COD	-310.6	-292.1	-43.9	-289.5	18.4	296.5
	(812.4)	(406.0)	(125.9)	(399.5)	(81.3)	(414.9)
Combo	6623.4***	1847.9***	583.1***	4418.5***	-5.18	-220.9
	(822.4)	(411.0)	(127.4)	(404.4)	(82.3)	(420.1)
N. of obs.	350	350	350	350	350	350
Mean control	5917.4	2083.5	274.8	2745.5	180.8	675.0
Combo-COD-CG	1364.9	593.0	-199.1	1668.4***	20.4	-717.8
p-value	0.27	0.33	0.29	0.0059	0.87	0.25
Panel B: Year 2						
CG	4958.1***	2213.0***	212.5	2400.2***	57.8	159.0
	(592.6)	(358.9)	(174.3)	(257.7)	(57.9)	(277.3)
COD	-274.3	28.0	-37.8	-253.4	-31.3	24.2
	(592.9)	(356.7)	(174.1)	(256.1)	(57.9)	(277.4)
Combo	5104.0***	2189.7***	645.3***	2460.6***	31.3	-217.6
	(600.2)	(361.1)	(176.7)	(259.3)	(58.6)	(280.8)
N. of obs.	350	349	349	349	350	350
Mean control	3669.3	1422.3	451.3	1314.3	96.6	414.2
Combo-COD-CG	420.2	-51.4	470.6*	313.8	4.92	-400.8
p-value	0.64	0.92	0.074	0.42	0.96	0.34

Table 6: Substitution from other sources

	Total	Government CG	Government Other	Local Government	NGOs	Parents	Other
Panel A: Year 1							
CG	725.8	749.6*	269.0	-137.5	0.35	-118.1	-34.5
	(615.6)	(387.1)	(254.3)	(275.4)	(9.02)	(211.8)	(163.0)
COD	-67.0	-17.0	246.8	115.3	-15.0*	-524.8**	103.4
	(615.0)	(387.0)	(254.2)	(277.5)	(9.03)	(211.7)	(163.0)
Combo	217.8	431.0	465.7*	-244.0	-10.6	-267.7	-159.2
	(623.0)	(392.0)	(257.3)	(278.9)	(9.16)	(214.4)	(165.1)
N. of obs.	350	350	350	350	350	350	350
Mean control	6095.8	4439.0	40.1	366.6	7.82	1084.1	158.3
Combo-COD-CG	-440.9	-301.6	-50.1	-221.8	4.11	375.1	-228.2
p-value	0.63	0.60	0.90	0.59	0.76	0.24	0.35
Panel B: Year 2							
CG	-1858.7*	-774.4*	-96.6	-10.4	158.6	-1087.3	-26.8
	(1037.4)	(404.5)	(349.4)	(35.5)	(116.9)	(919.2)	(69.5)
COD	-1006.5	-639.8	-122.1	44.9	-13.9	-348.5	74.4
	(1041.3)	(406.3)	(350.9)	(35.9)	(117.5)	(923.0)	(69.8)
Combo	-1069.6	-665.7	435.1	-20.6	129.0	-1180.5	186.1***
	(1049.8)	(409.6)	(353.6)	(35.9)	(118.6)	(930.0)	(70.4)
N. of obs.	349	349	349	349	349	349	349
Mean control	4759.2	2437.5	168.6	18.9	13.6	2046.9	73.7
Combo-COD-CG	1795.7	748.5	653.8	-55.1	-15.6	255.2	138.5
p-value	0.25	0.22	0.22	0.31	0.93	0.85	0.19

Table 7: Household expenditure

	Total Expenditure	Fees	Textbooks	Other books	Supplies	Uniforms	Tutoring	Transport	Others
Panel A: Year 1									
CG	-474.8	51.6	-67.3	23.5	311.3	404.6	-914.0	-216.0	-68.6
	(1821.7)	(474.0)	(92.4)	(75.8)	(309.2)	(848.7)	(954.1)	(224.2)	(246.0)
COD	1156.9	658.6	-33.0	-98.0*	381.9	286.6	-79.2	317.7	-277.6
	(1752.7)	(535.9)	(125.1)	(51.7)	(265.1)	(785.3)	(1039.3)	(500.8)	(274.4)
Combo	-2033.4	-821.8	131.0	117.5	-659.7	-1361.5	182.0	67.9	311.2
	(2568.4)	(781.1)	(154.6)	(104.7)	(471.9)	(1269.8)	(1465.0)	(688.0)	(409.4)
N. of obs.	1695	1695	1695	1695	1695	1695	1695	1695	1695
Mean control	26296.5	3241.0	268.5	138.4	4992.3	11286.7	4603.0	239.3	1527.3
Combo-COD-CG	-2715.4	-1532.1	231.3	192.0	-1352.8	-2052.6	1175.2	-33.8	657.5
p-value (H_0=Combo-COD-CG=0)	0.61	0.31	0.46	0.34	0.12	0.40	0.70	0.98	0.41
Combo-CG	-1558.5	-873.5	198.3	94.0	-971.0	-1766.1	1095.9	283.9	379.9
p-value (H_0=Combo-CG=0)	0.70	0.44	0.37	0.58	0.17	0.36	0.62	0.71	0.52
Panel B: Year 2									
CG	-3202.6	-1070.5	-222.4*	82.0	-172.5	-607.7	-524.8	-189.5	-497.1**
	(2213.0)	(771.4)	(118.2)	(93.4)	(278.0)	(710.6)	(874.5)	(459.1)	(249.7)
COD	1416.0	-537.5	-89.6	43.7	548.2*	630.1	1289.4	-500.6	32.2
	(2256.2)	(890.4)	(114.7)	(74.2)	(320.8)	(683.7)	(1010.6)	(398.8)	(222.4)
Combo	392.4	947.7	149.3	1.18	-258.8	190.3	-1107.3	12.2	457.7
	(2962.4)	(972.4)	(164.2)	(135.8)	(440.7)	(1016.3)	(1516.8)	(460.3)	(332.7)
N. of obs.	3318	3318	3318	3318	3318	3318	3318	3318	3318
Mean control	27466.6	2775.4	433.3	139.8	4155.0	14394.5	3161.9	427.3	1979.4
Combo-COD-CG	2179.0	2555.8	461.4	-124.5	-634.5	167.9	-1871.9	702.2	922.7
p-value (H_0=Combo-COD-CG=0)	0.74	0.29	0.19	0.62	0.48	0.94	0.51	0.58	0.19
Combo-CG	3594.9	2018.3	371.8	-80.8	-86.3	798.0	-582.5	201.7	954.8
p-value (H_0=Combo-CG=0)	0.45	0.21	0.16	0.70	0.89	0.62	0.79	0.82	0.082

Table 8: Test Scores

	Year 1				Year 2			
	Math	Swahili	English	Average	Math	Swahili	English	Average
CG	-0.051	0.0093	-0.021	-0.025	-0.0035	-0.021	0.0066	-0.0069
	(0.040)	(0.041)	(0.037)	(0.037)	(0.049)	(0.051)	(0.043)	(0.045)
COD	0.044	0.059	0.071*	0.069*	0.071	0.016	0.025	0.044
	(0.039)	(0.039)	(0.041)	(0.036)	(0.046)	(0.048)	(0.039)	(0.042)
Combo	0.10**	0.12***	0.078*	0.12***	0.19***	0.20***	0.096*	0.19***
	(0.045)	(0.040)	(0.043)	(0.040)	(0.047)	(0.045)	(0.051)	(0.045)
N. of obs.	8774	8774	8774	8774	9083	9083	9083	9083
Combo-COD-CG	0.11*	0.055	0.028	0.078	0.12*	0.20***	0.064	0.15**
p-value (H_0:Combo-COD-CG=0)	0.068	0.36	0.66	0.17	0.088	0.0051	0.34	0.019

Clustered standard errors, by school, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Effect on other grades and subjects

	Science		Grade 7 National Exam 2013		
	Year 1	Year 2	Pass rate	Average score	Test takers
CG	-0.0018	-0.048	-0.022	-1.95	2.20
	(0.054)	(0.055)	(0.025)	(1.80)	(3.04)
COD	-0.0079	-0.065	-0.014	-0.85	2.50
	(0.050)	(0.052)	(0.025)	(2.03)	(2.99)
Combo	0.058	0.050	0.011	0.69	3.11
	(0.052)	(0.050)	(0.026)	(2.14)	(3.08)
N. of obs.	8774	9083	337	337	337
Mean control group	-5.1e-09	7.1e-10	0.49	101.0	74.0
Combo-COD-CG	0.068	0.16**	0.047	3.49	-1.59
p-value (H_0:Combo-COD-CG=0)	0.38	0.041	0.23	0.25	0.73

Clustered standard errors, by school, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Focal years tutoring, tests and remedial

	Remedial	Tests	Tutoring
Panel A: Year 1			
CG	-0.033	0.69	0.012
	(0.031)	(0.81)	(0.034)
COD	0.0024	3.02***	0.066**
	(0.029)	(0.94)	(0.031)
Combo	0.034	0.24	0.040
	(0.029)	(0.80)	(0.031)
N. of obs.	1143	1133	1143
Mean of Dep. Var.	0.88	9.51	0.11
Combo-COD-CG	0.064	-3.47	-0.038
p-value	0.15	0.0047***	0.44
Panel B: Year 2			
CG	-0.0077	-1.05	-0.00086
	(0.058)	(1.31)	(0.024)
COD	-0.12**	0.042	0.025
	(0.055)	(1.04)	(0.032)
Combo	0.048	0.021	0.060*
	(0.044)	(0.92)	(0.034)
N. of obs.	965	960	965
Mean of Dep. Var.	0.77	9.84	0.071
Combo-COD-CG	0.18	1.03	0.036
p-value	0.031**	0.55	0.48
School Charact.	Yes	Yes	Yes
Teacher Charact.	Yes	Yes	Yes
District F.E.	Yes	Yes	Yes

Table 11: Time Use

	time preparing class	time teaching	time extra classes	time socializing colleague	Time (hrs) spend at school
Panel A: Year 1					
CG	-1.75	3.33	5.00	-0.23	0.054
	(3.27)	(5.63)	(3.29)	(2.14)	(0.12)
COD	2.14	-5.97	5.87*	2.24	0.12
	(2.97)	(5.91)	(3.40)	(2.24)	(0.10)
Combo	-0.64	2.32	5.34*	1.71	0.032
	(3.07)	(5.77)	(3.07)	(2.38)	(0.12)
Combo-COD-CG	-1.03	4.96	-5.52	-0.31	-0.14
p-value	0.84	0.56	0.29	0.93	0.42
Panel B: Year 2					
CG	3.14	-2.43	1.20	1.70	0.047
	(3.28)	(7.87)	(3.74)	(3.60)	(0.17)
COD	1.24	-19.0**	2.72	3.26	0.23
	(3.37)	(7.76)	(4.16)	(5.14)	(0.14)
Combo	-0.039	5.14	1.30	2.40	0.074
	(3.16)	(7.57)	(3.53)	(3.58)	(0.14)
Combo-COD-CG	-4.42	26.6	-2.62	-2.56	-0.20
p-value	0.37	0.018**	0.67	0.72	0.39

Table 12: Heterogeneity by student characteristics

	Gender	Age	Lag test score
Panel A: Math			
CG*Covariate	0.015	0.034	-0.021
	(0.065)	(0.022)	(0.041)
COD*Covariate	0.0082	-0.0077	-0.015
	(0.059)	(0.022)	(0.041)
Combo*Covariate	-0.061	-0.0097	-0.00027
	(0.059)	(0.022)	(0.040)
Covariate	0.061*	0.037***	0.40***
	(0.034)	(0.014)	(0.027)
Panel A: Swahili			
CG*Covariate	-0.0010	0.012	-0.0011
	(0.057)	(0.021)	(0.040)
COD*Covariate	-0.059	-0.030	-0.023
	(0.058)	(0.024)	(0.041)
Combo*Covariate	-0.088	-0.042**	-0.059
	(0.061)	(0.021)	(0.045)
Covariate	-0.046	0.023*	0.28***
	(0.033)	(0.012)	(0.029)
Panel A: English			
CG*Covariate	-0.028	0.0092	-0.051
	(0.047)	(0.018)	(0.067)
COD*Covariate	-0.084*	-0.0042	-0.043
	(0.048)	(0.021)	(0.060)
Combo*Covariate	-0.11**	-0.0091	0.017
	(0.049)	(0.019)	(0.079)
Covariate	0.030	-0.0084	0.19***
	(0.025)	(0.011)	(0.043)

Clustered standard errors, by school, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Effect of ignoring interaction effects: yr2 effects

	(1)	(2)
Panel A: Math		
CG	0.048	-0.0043
	(0.038)	(0.049)
COD	0.12***	0.070
	(0.036)	(0.046)
Combo		0.19***
		(0.047)
Panel B: Swahili		
CG	0.065*	-0.021
	(0.038)	(0.051)
COD	0.10***	0.014
	(0.037)	(0.048)
Combo		0.20***
		(0.045)
Panel C: English		
CG	0.034	0.0070
	(0.034)	(0.043)
COD	0.054	0.026
	(0.033)	(0.039)
Combo		0.097*
		(0.051)
Panel C: Average		
CG	0.059*	-0.0074
	(0.035)	(0.045)
COD	0.11***	0.044
	(0.034)	(0.042)
Combo		0.19***
		(0.046)

Clustered standard errors, by school, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Student present for low-stakes exam

	(1) Yr 1	(2) Yr 2
CG	0.011	0.019
	(0.013)	(0.012)
COD	0.015	0.0079
	(0.012)	(0.014)
Combo	-0.0092	0.013
	(0.012)	(0.013)
N. of obs.	9560	9390
Mean control	0.88	0.90
Combo-COD-CG	-0.035	-0.014
p-value	0.049**	0.48

Clustered standard errors, by school, in parenthesis
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$