Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

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

Isaac Mbiti (UVA)
Karthik Muralidharan (UCSD)
Youdi Schipper (Twaweza)

Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

Table of Contents

- Achieving universal primary education in developing countries is a top priority for global education policy (MDG)
- However, while developing countries have greatly improved school enrollment, learning outcomes are still poor

- Achieving universal primary education in developing countries is a top priority for global education policy (MDG)
- However, while developing countries have greatly improved school enrollment, learning outcomes are still poor
- In Tanzania, the net enrollment rate in primary school rose from 53% in 2000 to close to 90% in 2012
- But less than one third of grade 3 students could pass basic tests of second grade numeracy or literacy (Uwezo, 2013)

Background/Motivation

- Achieving universal primary education in developing countries is a top priority for global education policy (MDG)
- However, while developing countries have greatly improved school enrollment, learning outcomes are still poor
- In Tanzania, the net enrollment rate in primary school rose from 53% in 2000 to close to 90% in 2012
- But less than one third of grade 3 students could pass basic tests of second grade numeracy or literacy (Uwezo, 2013)
- Inadequate resources and poor teacher motivation are widely considered to be two key constraints to improving school quality in developing countries (including Tanzania)

- We study the effectiveness of policies aiming to alleviate both of these constraints with a large-scale RCT conducted across a representative sample of 350 schools in Tanzania

This paper

- We study the effectiveness of policies aiming to alleviate both of these constraints with a large-scale RCT conducted across a representative sample of 350 schools in Tanzania
- Interventions studied include:
    - A block capitation grant (CG) to schools
        - Large grant (doubled per-child non-teacher spending)
        - Subject to same guidelines as national CG program

- We study the effectiveness of policies aiming to alleviate both of these constraints with a large-scale RCT conducted across a representative sample of 350 schools in Tanzania
- Interventions studied include:
  - A block capitation grant (CG) to schools
    - Large grant (doubled per-child non-teacher spending)
    - Subject to same guidelines as national CG program
  - A Teacher performance pay program referred to as "Cash on Delivery" (CoD)
    - Paid to each teacher based on the number of students passing a basic test in each of 3 subjects (Math, Swahili, English)
    - NOT optimally designed (discuss reasons)

This paper

- We study the effectiveness of policies aiming to alleviate both of these constraints with a large-scale RCT conducted across a representative sample of 350 schools in Tanzania
- Interventions studied include:
  - A block capitation grant (CG) to schools
    - Large grant (doubled per-child non-teacher spending)
    - Subject to same guidelines as national CG program
  - A Teacher performance pay program referred to as "Cash on Delivery" (CoD)
    - Paid to each teacher based on the number of students passing a basic test in each of 3 subjects (Math, Swahili, English)
    - NOT optimally designed (discuss reasons)
  - A "Combination" (Combo) treatment arm that provided both interventions
    - Key innovation over existing work is that the design was explicitly powered to test for complementarities

- We find that
  - Increasing inputs (using capitation grants) has no impact on test scores

- We find that
  - Increasing inputs (using capitation grants) has no impact on test scores
  - This particular teacher incentive program had no impact on test scores on average (but some suggestive evidence of positive effects on students near the passing threshold)

- We find that
  - Increasing inputs (using capitation grants) has no impact on test scores
  - This particular teacher incentive program had no impact on test scores on average (but some suggestive evidence of positive effects on students near the passing threshold)
  - The combination of the two programs had a positive and significant impact on test scores
  - The combination is more than the sums of the parts (i.e., there are complementarities between the two)

- We find that
  - Increasing inputs (using capitation grants) has no impact on test scores
  - This particular teacher incentive program had no impact on test scores on average (but some suggestive evidence of positive effects on students near the passing threshold)
  - The combination of the two programs had a positive and significant impact on test scores
  - The combination is more than the sums of the parts (i.e., there are complementarities between the two)
- If complementarities matter, cross-cutting designs (which are widely used) could yield biased estimates if they ignore interactions

- **School Resources:** Glewwe, Kremer and Moulin (2009); Das, Dercon, Habyarimana, Krishnan, Muralidharan, and Sundararaman (2013); Sabarwal, Evans, and Marshak (2014); Blimpo and Evans (2011) ; Berry, Karlan, Pradhan, and Ratan (2012)

- **Teacher Incentives:** Glewwe, Holla, and Kremer (2009); Muralidharan and Sundararaman (2011); Muralidharan (2012); Duflo, Hanna, and Ryan (2012)

- **Estimation of Education Production Functions:** Todd and Wolpin (2003); Kremer (2003)

- **Complementarities in education production:** Muralidharan and Sundararaman (2011); Muralidharan 2012; No experimental study to date

Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

Table of Contents

Design

Figure: Districts in Tanzania from which schools are selected
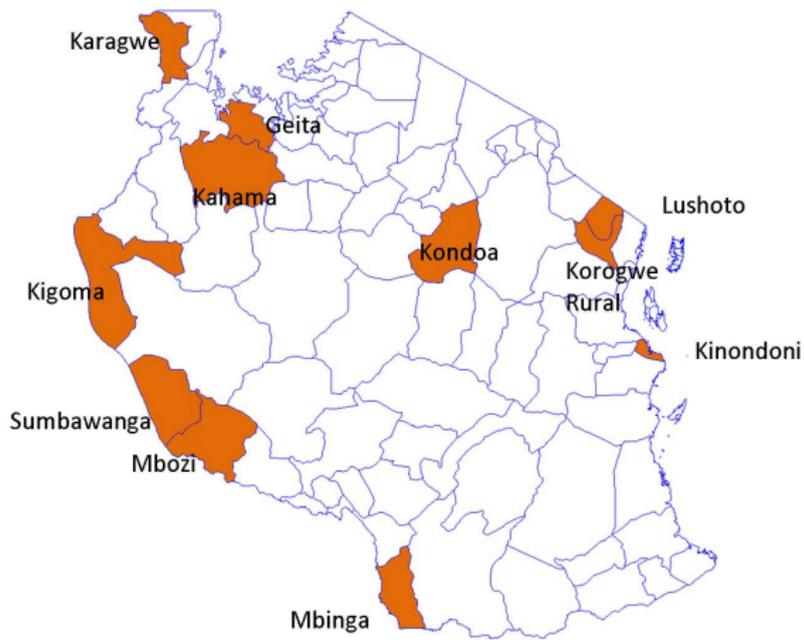
Implemented across a representative sample of 350 schools across 10 districts in Tanzania.

		CoD	
		Yes	No
CG	Yes	70	70
	No	70	140

Capitation grant (CG)

According to policy, the government is to disburse a capitation grant of TZS 10,000 per primary school pupil per year to school

- School committees may use however they like
- 85 Billion TZS ($\sim$ \$40M) per year
- Funds transferred to district councils, and not directly to schools
- On average about 4,700 TZS reached primary schools in recent years: "Considerable leakage" (World Bank, 2004)
    - On average 11 students share a single textbook!
    - 86% of schools do not have electricity
    - 60% of schools do not have a single classroom with a desk/chair for the teacher

Capitation grant (CG)

- Transfer capitation grants (CG) to reach schools in full (as per current government plan at TZS 10,000 per pupil/per year)
- The grant amount is about 3 times the mean pre-treatment school expenditure per student (excluding teachers salaries)
- Fund can be used to: buy books, booklets, small repairs, etc. but not for salaries or major construction
- Accountability as per existing government policy
- School leadership and communities informed on the intervention

Cash on Delivery (CoD)

- For every child in Grades 1, 2 and 3 who passes the proven Uwezo literacy (English and Kiswahili) and numeracy (Mathematics) assessment at the end of the school year, the child's teacher will be paid TZS 5,000 per subject the child passes (or up to TZS. 15,000 per each child who is able to pass all three literacy and numeracy tests)
- Head teachers will be paid TZS 1,000 per subject the child passes (or up to TZS. 3,000 per each child who is able to pass all three literacy and numeracy tests)
- Rewards absolute levels of learning, not gains
- Simple to implement and scale up
- Initially, low levels of learning (82% below the passing threshold for Swahili, 97% for English and 93% for Math)

- Give schools both a capitation grant (CG) and the Cash on Delivery (CoD).
- Adequately powered to test for complementarities

Research activities

Low-stakes exam (y1)
(Nov-Dec)

Baseline (y0)
(Feb/13)

Low-stakes exam (y2)
(Nov-Dec)

Information
sessions
(Mar-Apr)

Information
sessions
(Jul/13)

High-stakes
exam(Dec/13)

Information
sessions
(May/14)

Information
sessions
(Aug/14)

High-stakes
exam (Nov/14)
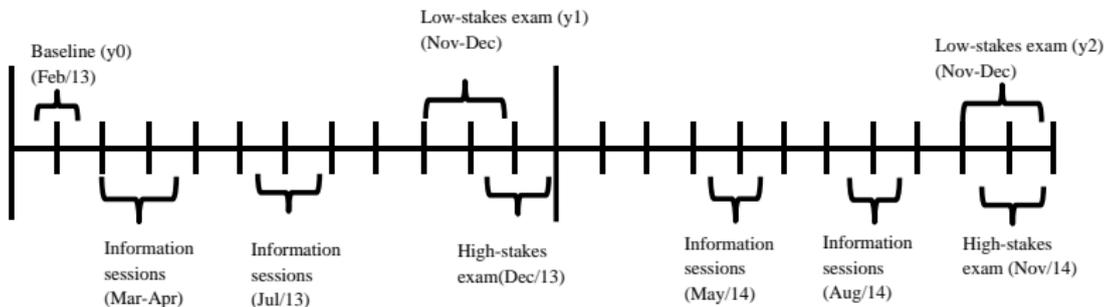
Intervention activities

- School data: facilities, expenditure, enrollment, etc.
- Teacher data: socio-demographic characteristics, qualifications, experience, time use, etc.
- Student data: We test 10 students from each focal grade (grades 1, 2 and 3), in all three focal subjects (Math, English and Swahili) and in Science.
    - This is a low stakes exam different from the "intervention" exam.
- Household data (for 1/3 of students): household and dwelling socio-demographic characteristics, household expenditure in education, parents engagement in children's education, etc.

- No difference in baseline characteristics by ▸ student , ▸ household , ▸ school , or ▸ teachers
- No differential attrition in test-taking sample ▸ attrition

Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

Table of Contents

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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

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

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

Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

Table of Contents

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: 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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

"Conducting a series of evaluations in the same area allows substantial cost savings. Once staff are trained, they can work on multiple projects. 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 AER, 2003)

Early wave of experiments

- Low budgets
- Cross-cut designs were very popular
- Interaction effects were considered a second order issue

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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

Table of Contents

- No systematic evidence of more teacher effort (time teaching, time spend at schools, etc.) ▸ Time usage
- No systematic evidence of change in pedagogy (tests, homework, tutoring, etc.) ▸ Pedagogy

Introduction
Design and data
Results - Expenditure
Results - Test scores
Mechanisms
Heterogeneity

Table of Contents

- No heterogeneity by baseline characteristics (▸ Student characteristics)
- No heterogeneity by baseline test score: parametric and non-parametric (▸ Math , ▸ Swahili , ▸ English)

- No differential impact for CG (not surprising - zero on average)
- Suggestive evidence of larger effects in incentive treatments closer to the passing threshold (stronger for Combo)
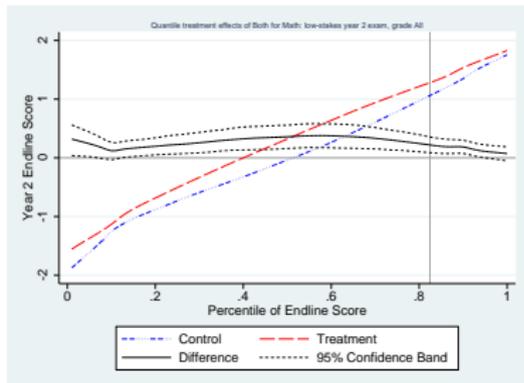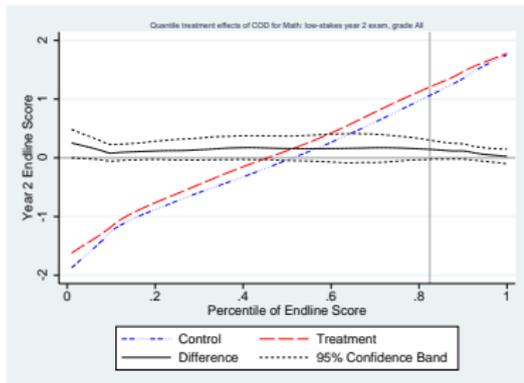
Figure: Effect of Cash on Delivery (left) and Combo (right) on Math test scores by endline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.
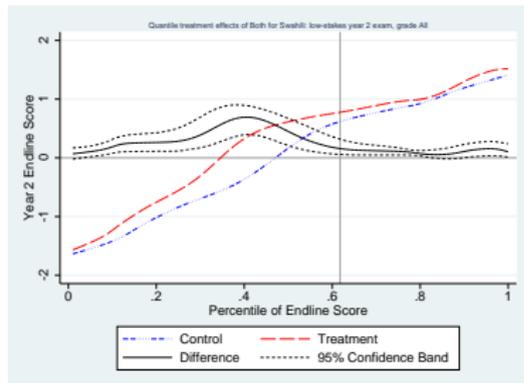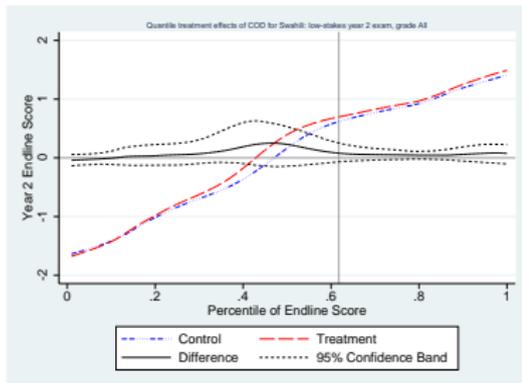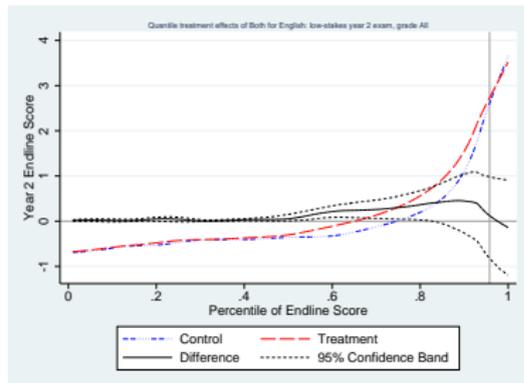
Figure: Effect of Cash on Delivery (left) and Combo (right) on Swahili test scores by endline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.
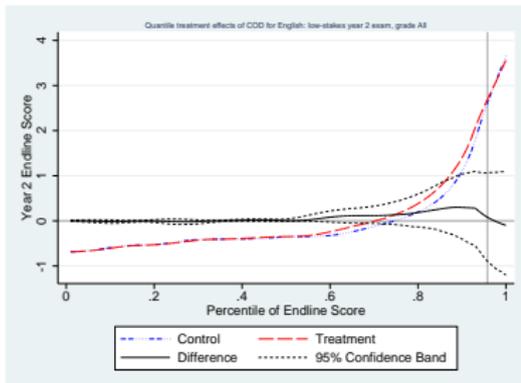
Figure: Effect of Cash on Delivery (left) and Combo (right) on English test scores by endline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.

- No impact on test scores of additional school grants
- Insignificant positive impacts of teacher bonuses on average, but find suggestive evidence of threshold effects
- Test scores in schools with both programs were significantly higher than those in control schools, and we find strong evidence of complementary between inputs and incentives
- Effectiveness of teacher performance pay programs are likely to crucially depend on how they are designed
- Cross-cutting experimental designs that ignore interactions may yield biased estimates of the main treatment effects of interest

- Admin: administrative cost (including staff wages), rent and utilities, and general maintenance and repairs
- Student: food, scholarships and utilities (notebooks, pens, etc.)
- Teaching aid: classroom furnishings, textbooks, maps, charts, blackboards, practice exams, etc.
- Teachers: salaries, bonuses and teacher training

	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

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

	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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$

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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
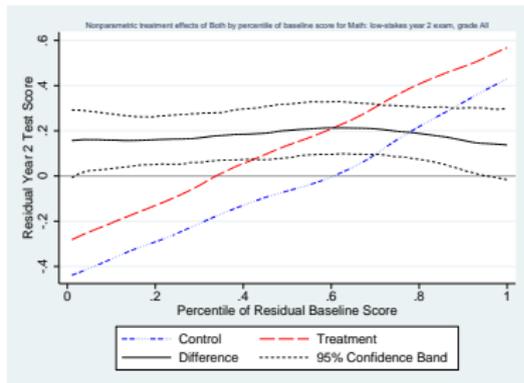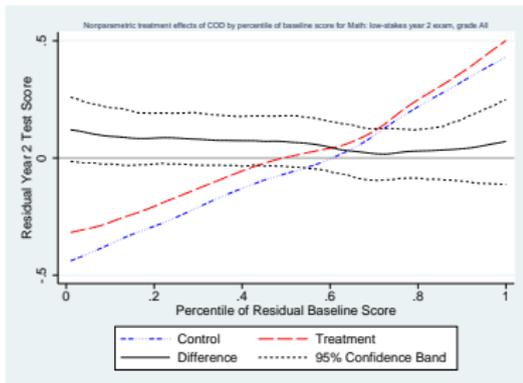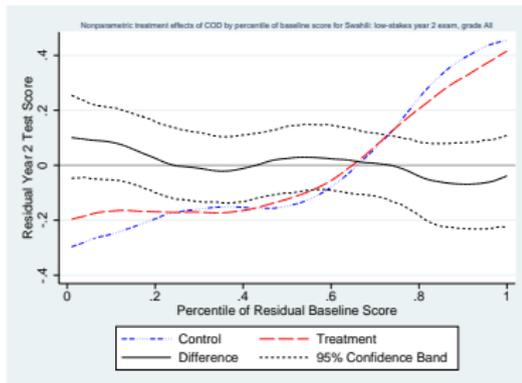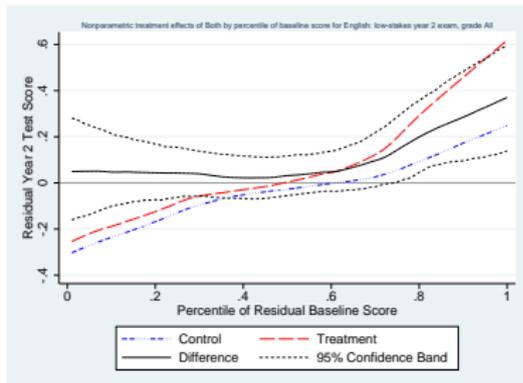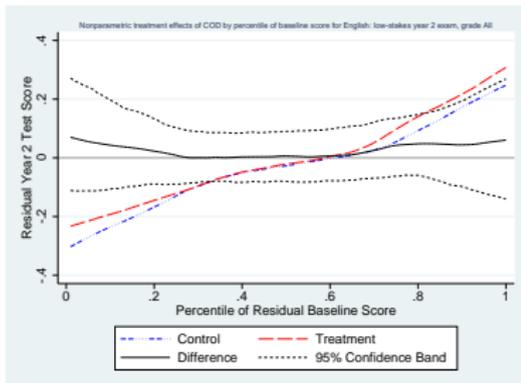
Figure: Effect of Cash on Delivery (left) and Combo (right) on Math test scores by baseline score using a lowess regression. Standard errors are calculated using bootstrapping and clustered at the school level.

Figure: 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: 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.