RISE Working Paper 20/047
September 2020

Teacher Performance-Based Incentives and Learning Inequality

Deon Filmer, James Habyarimana, Shwetlena Sabarwal

Abstract

This study evaluates the impacts of low-cost, performance-based incentives in Tanzanian secondary schools. Results from a two-phase randomized trial show that incentives for teachers led to modest average improvements in student achievement across different subjects. Further, withdrawing incentives did not lead to a "discouragement effect" (once incentives were withdrawn, student performance did not fall below pre-baseline levels). Rather, impacts on learning were sustained beyond the intervention period. However, these incentives may have exacerbated learning inequality within and across schools. Increases in learning were concentrated among initially better-performing schools and students. At the same time, learning outcomes may have decreased for schools and students that were lower performing at baseline. Finally, the study finds that incentivizing students without simultaneously incentivizing teachers did not produce observable learning gains.

JEL Classifications: I21; I28; J45

Keywords: student achievement; teacher incentives; student incentives; learning outcomes; inequality

Teacher Performance-Based Incentives and Learning Inequality

Deon Filmer
The World Bank

James Habyarimana
Georgetown University

Shwetlena Sabarwal
The World Bank

Acknowledgements:

We thank the staff of the Tanzania Ministry of Education, Science and Technology for extensive discussions on program design and implementation. We are grateful for comments from Samer Al Samarrai, Alaka Holla, Tobias Linden, Ezequiel Molina, and participants at the 2019 Research on Improving Systems of Education (RISE) conference. Tummaniel Ngowi, Sana Khan, Tiffany Tong, and Ravinder Gera provided excellent field-based support. Avichal Mahajan, Ali Hamza, Olivia Bisel, Frances Chen, Rachel Leeds, and Chinmaya Holla were superb research assistants. Financial support from the Strategic Impact Evaluation Fund, Multi-Donor Education and Skills Fund, Strategic Research Program, and RISE is gratefully acknowledged.

This is one of a series of working papers from "RISE"—the large-scale education systems research programme supported by funding from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), the Australian Government's Department of Foreign Affairs and Trade (DFAT), and the Bill and Melinda Gates Foundation. The Programme is managed and implemented through a partnership between Oxford Policy Management and the Blavatnik School of Government at the University of Oxford.

Please cite this paper as:
Filmer, D., Habyarimana, J. and Sabarwal, S. 2020. Teacher Performance-Based Incentives and Learning Inequality. RISE Working Paper Series. 20/047. https://doi.org/10.35489/BSGRISEWP_2020/047.

1. Introduction

Given the structure of their contracts and their political influence, teachers, especially in low- and middle-income countries, are only weakly extrinsically incentivized to exert effort towards student learning (Hanushek 1996, Lockheed and Verspoor 1991, World Bank 2004, 2018). Performance-based incentives have emerged as a potential way to align teacher effort towards student learning. This is motivated by two features of the education sector. First, observable teacher characteristics such as experience and qualifications, which are the main determinants of salaries in most school systems, are weak predictors of teacher effectiveness in producing student learning (Rivkin, Hanushek and Kain 2005; Aaronson, Barrow and Sander 2007). Second, using direct personal incentives might be better than intervening in the educational process directly because policy makers may not know the best means of improving education given the heterogeneous attributes of students and teachers. Providing incentives tied to student performance allows schools and teachers to choose the best means to improve performance given their circumstances.

Data, albeit from primary schools, suggest that there are significant margins to increase teacher effort in Tanzania. In 2014, students were being taught for about 2 hours and 47 minutes per day out of 5 hours and 56 minutes of scheduled teaching time (Martin and Wane 2016). There were no major differences in learning time across urban and rural schools. Like in many other Sub-Saharan African countries, participation rates at the secondary school level have increased dramatically in Tanzania, for example the lower secondary gross enrollment rate increased from 20 percent in 2000 to close to 60 percent in 2016 (Bashir et al. 2018). At the same time there have been persistent concerns with poor learning outcomes at both the primary and secondary levels, which have, in turn, been linked to low levels of subject knowledge and low effort among teachers (Bold et al. 2017a; 2017b).

Our study's main goal is to assess whether addressing teacher effectiveness through fiscally and politically viable low-cost performance-based incentives can be sufficient to make a difference to student learning outcomes. We further investigate five aspects of such an intervention. These aspects emerged as areas of concern expressed by Tanzanian education policymakers[1] in the context of discussions around the

[1] The research team discussed the evaluation design with the Permanent Secretaries (PS) in Tanzania Ministry of Education and Vocational Training (MoEVT) (which is the antecedent of the current Ministry of Education, Science and Technology) and President's Office, Regional Administration and Local Government. The team worked closely with MOEVT focal points assigned by MoEVT PS.

potential benefits and drawbacks of introducing teacher incentives. First, we evaluate the impact of withdrawing the incentives with the goal of establishing whether one might find a discouragement effect. Second, we assess the durability of the impacts by following students beyond the intervention period. Third, we document heterogeneity in the impact of the teacher incentives. Fourth, we explore the pathways through which the effects might work. Last, we contrast incentives for teachers with those for students.

We study these questions through a two-phase randomized research design focusing on Grade 10 teachers and students in about 400 schools in Tanzania. Study schools were randomly assigned into one of a set of treatment arms in each phase of the program. In phase 1, schools were randomly assigned to either a control group, an arm implementing teacher incentives, or an arm implementing both teacher and student incentives. Teachers in schools assigned to the teacher incentive arm "competed" with teachers in approximately 6-12 other comparable schools—and teachers with the greatest value added in terms of student test scores received an award. Students assigned to the student incentive arm "competed" with students in the same school—and students with the greatest increase in terms of test scores received an award. In phase 2, student and teacher incentives programs were implemented but with a twist. First, teacher incentive schools in phase 1 were randomly split into schools that would implement teacher incentives for the second year and schools where those incentives were withdrawn. Second, the control group from phase 1 was randomly split into a pure control group and a student incentives-only group. So overall, schools were randomly assigned to either a control group, an arm implementing teacher incentives for a second year, an arm in which teacher incentives were discontinued, and an arm implementing student incentives alone. The design is summarized in Figure 1.[2]

Our main finding is that teacher incentives led to a modest average increase in student learning outcomes. In phase 1 these were only found when combined with student incentives, in phase two these were found for the simple teacher incentive intervention. The effect sizes we find range from 0.09 to 0.18 standard deviations (sd), depending on the phase and the test score (Mathematics, English, or Kiswahili) used to measure impacts. We do not find any evidence of a discouragement effect when teacher incentives were removed: in fact, the effects remain positive, although they are generally smaller and no longer statistically significant. Repeating the teacher incentives intervention in phase 2 resulted in similar impacts on student learning as in phase 1, suggesting that there was neither learning nor fatigue associated with the second round of teacher incentives. Next we show that learning impacts persisted. Specifically, when we follow-up with the phase 2 cohort of students more than one year after the intervention—when they are no

[2] In Tanzania, teachers are typically linked to grades (not student cohorts). Since, we target the same grades in Phases 1 and 2, by design, the same teachers were involved in each phase. However, teacher turnover in Tanzania is high, therefore there was some change in the teacher groups between phases. Our school level estimates suggest that two thirds of teachers remain in the school across both phases of the study.

longer the students of an incentivized teacher and when all incentives have been completely withdrawn from the schools—we find effect sizes to be statistically significant and on the order of 0.1 sd. The last of the average effects we document are that learning outcomes did not increase as a result of student incentives when these incentives were implemented on their own (phase 2). When combined with teacher incentives (phase 1) they were more generally associated with increases in learning outcomes.

Beyond establishing modest average impacts, our second main finding is that there was substantial heterogeneity in learning impacts from the teacher incentives. We find this both across schools—learning impacts were concentrated in schools that had higher baseline test scores, and across students, learning impacts were largest for students in the higher (conditional) quantiles of the test score distribution—with effect sizes reaching as high as 0.4 standard deviation for some groups of students. At the same time, the impact in initially poorly performing schools was close to zero and sometimes negative for students in the lower (conditional) quantiles of the test score distribution (although not always statistically significantly so). Both dimensions of heterogeneity (across and within schools) go in the direction of exacerbating inequalities. This suggests that while, in this case, teacher incentives may have had a positive average impact, the benefits to students in terms of learning were very unequally distributed.[3] We also find that, in phase 2, student incentives implemented on their own had a positive impact on student test scores in schools that had higher baseline scores (but not in other schools).

Our paper contributes to two main literatures, namely those on teacher and student incentives. Empirical evidence on the effectiveness of teacher incentives is mixed.[4] Positive impacts on learning, with effects typically on the order of between 0.15 to 0.3 sd, have been found in randomized evaluations in a number of low-income countries (Glewwe et al. 2010 in Kenya; Loyalka et al. 2019 in China; Gilligan et al. 2019 in Uganda, Duflo et al. 2012, and Muralidharan and Sundararaman 2011a, 2011b in India).[5] In primary schools in Tanzania, Mbiti et al. (2019) found that teacher incentives led to a 0.21 sd increase in student scores after two years on the test used to set the awards but no impact on a separately administered test; when combined with (non-incentive-based) school grants, they found statistically significant impacts of 0.36 sd and 0.23 sd respectively on these tests.[6] In contrast, no impacts were found in an intervention that linked bonuses to student exam scores in Pakistan (Barrera-Osorio and Raju 2017). The finding of no

[3] The observed teacher incentives or incentives withdrawn impacts in phase 2 are the average effects over groups that had received different treatments in phase 1 (teacher incentives only and teacher and student incentives). The study design is not powered to detect potential differences by phase 1 assignment.

[4] See Glewwe and Muralidharan (2016) for a recent review of the impact of teacher incentives for developing countries.

[5] Positive and significant impacts are also seen in quasi-experimental studies in Israel (Lavy 2002, 2009).

[6] In an evaluation of a pay for performance program in Austin, Texas, Balch and Springer (2015) find that the program is associated with positive student test score gains in both math and reading during the initial year of implementation but do not find any additional growth in the second year.

impacts has been documented in a number of high-income settings (Glazerman and Seifullah 2012; Springer et al. 2012 in the United States), and some have even found negative impacts (Martins 2009 in Portugal; Fryer 2013 in New York City, Atkinson et al. 2009 in England).

Recent meta-studies of interventions aimed at increasing learning outcomes in primary schools in low- and middle-income countries conclude that teachers do indeed respond to incentives (Evans and Popova 2016, McEwan 2015), especially in low-accountability settings (Ganimian and Murnane 2016). Examining the mixed evidence more closely suggests that impacts are likely to depend on context, especially on the existing margins for improving teacher effort. Incentives are likely to be more effective when teachers can determine and implement low-cost actions to improve student learning. This is likely to be the case in developing countries, where teacher absenteeism is high and time on task is low (Chaudhury et al. 2006, Bold et al. 2017b). Impacts also depend on the details of the design features of the incentive program, for example whether the incentives are for groups versus individuals, or whether rewards are financial versus non-financial (Bruns, Filmer and Patrinos 2011) or how performance is measured (Barlevy and Neal 2012).

A related literature focuses on perverse impacts of teacher incentives. This includes "teaching to the test" which was documented in Kenya (Glewwe et al. 2010)—although in another context (India) positive impacts spilled over to non-incentivized subjects (Muralidharan and Sundararaman 2011a). Much of this literature has focused on cheating, which has been documented in both middle- and high-income contexts (Behrman et al. 2015 in Mexico; Jacob and Levitt 2003 in the United States). Neal (2008) discusses a variety of ways in which schemes that link teacher incentives to student achievement could potentially be corrupted (also see Cullen and Reback 2006, Jennings and Beveridge 2009, Figlio, 2005 and Jacob 2005, 2007).

The literature on what happens when teacher incentives are withdrawn or on the heterogeneity in impacts is relatively thin. On the first, Jinnai (2016) found that in North Carolina (USA), where the state government first reduced and finally repealed its teacher incentive program, student achievement at the lowest-performing schools significantly decreased. On the second, Bacache-Beauvallet (2006) argues theoretically why incentives would increase inequality, while Chang et. al (2020) evaluate a teacher incentive program that explicitly rewarded performance increases of initially poorer-performing students, and indeed found larger impacts for these students.

The second literature we contribute to is that on student incentives where empirical evidence from low- or middle-income settings is thin and inconclusive. Positive impacts have been documented in Kenya

(Kremer et al. 2007) and Benin (Blimpo 2014).[7] On the other hand, Li et al. (2010) found that a cash incentive linked to performance led to no measurable impacts on learning outcomes in China.[8] Evidence on student incentives programs from high-income countries is likewise mixed, finding no impacts overall (Fryer 2011 in three US cities); inconsistent results by gender of the student (Angrist and Lavy 2009 in Israel; Angrist et al. 2009 in Canada) or by subject (Bettinger 2012 in Ohio); positive impacts for students close to meeting the achievement standard in the United States (Levitt et al. 2016). Jackson (2010) found that combining teacher and student incentives had a significant impact on exam performance in Texas (USA).

There is only one study, to our knowledge, that directly compares teacher and student incentives (Behrman et al. 2015). The study experimentally contrasts the impact of three different performance incentive schemes in Mexican high schools and found that the largest impacts were for a "complete" package of individual and group incentives to students, teachers and school administrators, smaller impacts for individual student incentives only, and no impacts for the individual teacher incentives arm.

Our paper is organized as follows. Section 2 provides background information on the incentive program and the context of the study, experimental design, data, and estimation strategy; Section 3 presents the main results. Section 4 concludes.

2. Context, Experimental Design, Data, and Estimation Strategy

The program we evaluate was implemented across a sample of lower secondary schools in three regions of Tanzania.[9] The intervention focuses on students in Grade 10 (Form 3, the third of four years of lower secondary school). The context is one in which teacher compensation and promotion are largely unrelated to performance in terms of student learning outcomes. Procedures for teacher recruitment and firing are

[7] A study from Cambodia showed that transfers based on baseline performance (not linked to subsequent effort) had positive impacts on learning (Barrera-Osorio and Filmer 2016).

[8] In an experiment in non-formal schools in Indian slums, a reward scheme for attending a target number of school days increased average attendance when the scheme was in place. However, among students with low baseline attendance, the incentive lowered post-incentive attendance and test scores—pointing to unintended negative consequences of the intervention (Visaria et al. 2016).

[9] These are Lindi, Morogoro, and Shingyanga. Due to re-districting in Shinyanga during the course of the study, some schools were re-classified into Geita and Simiyu. However, for the purposes of this study, schools retain their 2012 regional classification. These three regions represented roughly 11 percent of Tanzania's population in 2012. They were selected based on government guidance to reflect intermediate levels of socio-economic development across Tanzania. In the 2015 HDI index ranking of Tanzania's 26 regions (with Rank 1 being best performing and Rank 26 being the worst performing), Lindi ranked 12, Morogoro 15, and Shinyanga 19. They also reflect high vs. low levels of private school penetration. In Morogoro, 21 percent of its total secondary schools were private in 2012 and in Shinyanga and Lindi this share was 8 percent and 6 percent, respectively.

complicated and lengthy. In addition, monitoring and supervision of teacher performance by school inspectors is weak and school management committees have little or no oversight over teachers (United Republic of Tanzania 2008). While improvements in the management and supervision of teachers are long-run goals, the objective of this study was to determine whether, in the short run, providing performance-based incentives to teachers and students could boost learning outcomes. Because fiscal constraints and political pressures would make it hard to scale up more high-powered incentives, the decision was made to use a relatively low-stakes and low-cost approach to incentives.[10]

2.1 Experimental design

This research was initiated in 2013 and comprises of two phases: phase 1 covered the 2013 academic year (January-December) while phase 2 covered the 2014 academic year. Around 420 secondary schools (public and private) in three regions of Tanzania—one in the north-western part of the country (Shinyanaga) and two in the south-east (Lindi and Morogoro)—were sampled. The interventions were restricted to Grade 10 students and teachers and focused on student performance in Math, English, and Kiswahili. Because most students would have passed from Grade 10 to 11 between 2013 and 2014, the cohort of students in phase 1 is different from that in phase 2.

The full experimental two-phase design is summarized in Figure 1. In phase 1, schools were randomly assigned to one of two treatment arms or a control arm. In the first treatment arm teachers were eligible to receive performance incentives (TI), in the second treatment arm both teachers and students were eligible to receive incentives (TI+SI). The control group received nothing. In phase 2, assignments to treatment groups were based on innovations to phase 1 treatments, to best answer policy maker's questions. Specifically, phase 1 control was randomly assigned to one of two phase 2 arms: a pure control and a student-incentives only arm (SI). In addition, phase 1 TI and TI+SI arms (together) were randomly divided into two phase 2 arms. In the first of these, teacher incentives were offered again (TI) and in the second, the teacher incentive program was discontinued (TI withdrawn). Student incentives were offered in neither of these groups.

The teacher incentive program was structured as follows. Schools were divided into relatively homogeneous groups within which teachers would "compete," based on geographic sub-region and performance at baseline (6-12 schools per group). The objective of restricting the set of comparators to similar schools was to help ensure that all teachers believed that they had a reasonable chance of winning

[10] See the discussion in Grindle (2004) on political pressures. An additional reason to use non-financial incentives is because these may crowd out intrinsic motivation (List, Livingston and Neckermann 2018; also see discussion in Benabou and Tirole 2003).

an award. Awards for teachers were based on annual average gain in the score of the teacher's students between baseline and endline tests within the same phase.[11] Gains were expressed as percentage-point increases in the score on the test, which was marked out of 100. The curriculum-based tests were designed specifically for this intervention and were administered at the beginning and end of the school year. Within each group of 6-12 schools, the top three teachers in each of Math, English, and Kiswahili received an in-kind award. These were distributed at region-wide public ceremonies chaired by high-ranking district and regional officers. Awards included smart phones, book vouchers, certificates, and medals with an overall value of roughly $190 for first-place teachers, $130 for second-place teachers, and $110 for third-place teachers.[12]

The student incentives were designed as within-school competitions. Within each school, students achieving the highest annual gains in the average score on the tests received an award. Again, gains were calculated in percentage points on the study-specific standardized curriculum-based test, which was marked out of 100. For each eligible school assigned to this arm, the top three students in Math, English, and Kiswahili were awarded. Awards, which included book vouchers, certificates, and medals, were distributed at school-based ceremonies.

The total annual cost per year for the teacher incentive program was approximately US$ 442, and that for the student incentive program was approximately US$ 92. This is excluding student assessment costs under the assumption that the awards can be tied to regular system-level student assessments.[13]

2.2 Sampling

Sampling of schools was done using a complete list of public and private secondary schools in the three regions.[14] All wards with private schools were automatically included in the sample. After this step was completed, remaining wards were selected randomly, with the goal of reaching a sample size of over 400

---

[11] The announcement for the awards included a note that students who were absent on the day of the endline test would get a score of zero. Students who were absent at baseline but present at endline would receive the average baseline score for the school.

[12] Analysis of Tanzania's 2014 Labor Force Survey suggests that median monthly earnings of secondary school teachers are on the order of $376. This is reported as PPP$934 in Evans, Yuan, and Filmer (2020) based on a survey median of Tanzanian shillings 622,434. $376 is calculated on the basis of an exchange rate of 1653 shillings per dollar as reported in World Development Indicators for 2014.

[13] Intervention costs increase by approximately US$ 500 per school, per year if the cost of designing, administering, invigilating, marking, and data-entry of student assessments is factored-in.

[14] The sample for the study was constructed using complete secondary school lists for three regions in Tanzania – Lindi, Shinyanga, and Morgoro. The school list, disaggregated at the ward level, is derived by triangulating two administrative databases - EMIS 2011 and Inspectorate 2011 data. These data were augmented by a school mapping database created in 2013 as a part of the fieldwork for this research. We included all the schools for which we could find school performance data (Form 2 pass rates) for 2012. The final sample included 61 percent of all public secondary schools and 100 percent of all private secondary schools in the three regions.

schools and with the distribution of schools across regions being in proportion to their relative population size (using the 2002 census). The final sample consists of 62 schools in Lindi, 138 in Morogoro and 220 in Shinyanga.[15,16]

Schools were initially randomly assigned to different treatment groups using a three-step process. First, for each region, a natural category (region, district, or combination of districts) with about 60 schools was constructed.[17] Second, within each natural category, we generated a ranking of school quality using Form II pass rates (derived from Inspectorate Data) and then we created three strata using terciles of the pass rate ranking.[18] Third, random assignment was undertaken as follows: (i) within each tercile, half of the schools were randomly assigned to receive teacher incentives with other half as control; (ii) within the teacher incentives arm, half of the schools were randomly assigned to teacher incentives only and the other half to teacher incentive plus student incentives; and (iii) within the control arm, half of the schools were randomly assigned to receive another treatment which turned out to be unfeasible and these schools were not visited again in Phase 1, with the other half assigned to have continual data collection during phase 1. This produces a sample with phase 1 assignments of roughly 25 percent to teacher incentives only, 25 percent to both teacher and student incentives, 25 percent to unfollowed control, and 25 percent to followed control.

In phase 2, half of the schools in which teacher incentives had been implemented (either with or without student incentives) were randomly selected to continue with the scheme, and half of the schools were selected to discontinue it. In addition, the control group from phase 1 (including the unfollowed arm) was randomly split into a pure control group and a student-incentives only group. This produces a sample with phase 2 assignments of roughly 25 percent to teacher incentives (repeated), 25 percent to teacher incentives withdrawn, 25 percent to student incentives, and 25 percent to control.

[15] At the time that the sample was generated, a list of replacement schools was also generated from the population of schools. Schools were replaced for two reasons: (i) the school was no longer functioning or where duplicates were found; or (ii) the school either did not have a Form 3 cohort or had 10 Form 3 students or fewer.

[16] At the time of Phase 1 Baseline (November 2012), the full impact evaluation sample of 409 schools included 206 control schools. These 206 schools were then sub-divided into two groups - 105 pure control schools which were not visited again in Phase 1 and 101 control schools which were visited in Phase 1. Consequently, in Phase 1 midline and endline, the total number of schools visited is 304 (409 – 105). In Phase 2, the 105 pure control schools were added back to the evaluation sample. Between Phase 1 Baseline and Phase 2 Baseline, the school attrition was around 13 schools (409-396), due to school closures or school desire not to participate further in the impact evaluation. During Phase 2 another 3 schools dropped out of the evaluation. The school sample is smaller for Phase 2 midline (385) because of the timing (September 2014) when some schools were found closed for examination preparation holiday.

[17] For Lindi, there was no sub-division. The Morogoro region was divided into 3 such categories and Shinyanga was divided into 4

[18] When Form II pass rates were not available, the value was predicted by using 1) Form IV pass rates or 2) baseline results. When no data were available, the schools were assigned to tercile 2.

2.3 Data

Data collection began in November 2012 and lasted until September 2015 (Table 1). For each phase, three rounds of data were collected: baseline data at the start of the school year; midline data in the middle; and endline data at the end of the school year. Phase 1 baseline data were collected in November 2012-February 2013, midline in August-September 2013, and endline in November 2013. Phase 2 data were collected at corresponding times in 2014. At baseline and endline, students were tested and teachers and head-teachers were surveyed. At midline, attendance data were collected for both students and teachers during unannounced visits. Teachers were observed teaching a class during the midline visits and their behaviors recorded along a number of dimensions using an instrument with 21 items across the following domains: how teachers interact with students; the materials they used during instruction; use of the blackboard; and assignment and review of homework.[19] After phase 2, post-endline data were collected in August-September 2015 during which the phase 2 cohort of students was re-tested only in math. Test scores used in our empirical analysis are all normalized to have a mean of 0 and a standard deviation of 1 for the control group in the relevant phase.

Around 83 percent of the schools in our sample are public schools; 92 percent are O Level only (go up to Grade 11). At time of the Phase 1 baseline (2013), around 36 percent of schools had electricity and almost all schools (99 percent) collected school fees. Around 56 percent of the schools in our sample have less than 10 permanent teachers; 11 percent have more than 20 permanent teachers. The majority of head-teachers (85 percent) and teachers (79 percent) are male. Nearly 83 percent of teachers have more than secondary education and 90 percent have received some type of formal teacher training.

2.4 Estimation strategy

We divide our empirical investigation into four main parts. We start by analyzing average impacts on student test scores in phases 1 and 2. Given the random assignment study design, our estimation of treatment impacts is straightforward. All of the estimates described below are intent-to-treat. We regress endline test scores $Test_{i1\_e}$ for phase 1 and $Test_{i2\_e}$ for phase 2 on the set of indicators capturing random assignment to a study arm. We control for a set of indicators to capture the stratified random assignment

[19] For each phase, we conducted a separate exploratory factor analysis and retained the top four leading factors. In phase 1, the leading factors, which explain just under 38 percent of the variation across all items are shaped by items related to a) teacher-pupil interactions, b) use of materials such as maps, textbooks or equipment in instruction; c) use of the blackboard to copy lessons and illustrate examples, and d) assignment and review of homework. In phase 2, the four leading factors explain 36 percent of total variation and their dimensionality is explained by the same sets of items as the Phase 1 leading factors.

to arms (Region*PassTercile). We also adjust for an indicator that the school's baseline average test score exceeds 50 percent ($B50_i$). Our main specifications are therefore:

Phase 1: $Test_{i1\_e} = \beta_{1\_0} + \beta_{1\_1}*TI\_only_i + \beta_{1\_2}*TI\_SI_i + \delta_{1\_0}*B50_i + \Sigma_{rj}\kappa_{1\_rj}Region_r*PassTercile_j + \varepsilon_{1\_i}$ (1)

Phase 2: $Test_{i2\_e} = \beta_{2\_0} + \beta_{2\_1}*TI\_only_i + \beta_{2\_2}*TI\_withdrawn_i + \beta_{2\_3}*SI\_only_i + \delta_{1\_0}*B50_i + \Sigma_{rj}\kappa_{2\_rj}Region_r*PassTercile_j + \varepsilon_{2\_i}$ (2)

Next, we investigate heterogeneity in impacts using two approaches. First, in order to explore cross-school heterogeneity, we rely on machine learning to identify the covariates associated with treatment heterogeneity.[20] Guided by the machine learning results, we include an indicator variable for whether the baseline average score for the school was above 50 percent ($B50_{i1}$) and interact that indicator with the indicators for each study arm. Roughly 30 percent of schools are above that threshold. The specifications we estimate are therefore:

Phase 1: $Test_{i1\_e} = \beta_{1\_0} + \beta_{1\_1}*TI\_only_i + \beta_{1\_2}*TI\_SI_i + \delta_{1\_0}*B50_{i1} + \delta_{1\_1}*B50_{i1}*TI\_only_i + \delta_{1\_2}*B50_{i1}*TI\_SI_i + \Sigma_{rj}\kappa_{1\_rj}Region_{ri}*PassTercile_j + \varepsilon_{1\_i}$ (3)

Phase 2: $Test_{i2\_e} = \beta_{2\_0} + \beta_{2\_1}*TI\_only_i + \beta_{2\_2}*TI\_withdrawn_i + \beta_{2\_3}*SI\_only_i + \delta_{2\_0}*B50_{i1} + \delta_{2\_1}*B50_{i1}*TI\_only_i + \delta_{2\_2}*B50_{i1}*TI\_withdrawn_i + \delta_{2\_3}*B50_{i1}*SI\_only_i + \Sigma_{rj}\kappa_{2\_rj}Region_r*PassTercile_j + \varepsilon_{2\_i}$ (4)

Second, in order to explore cross-student heterogeneity, we re-estimate equations (1) and (2) using quantile regressions. Specifically, we estimate the impacts on students at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles of the distribution of $\varepsilon_{1/2\_i}$, namely the test score conditional on all the variables capturing the

[20] We use the *generalized random forests* machine learning algorithm (Athey and Wager 2019) to identify the sources of variation in conditional average treatment effects. Observations are clustered at the school level with a minimum cluster size of 20 students. Across the three treatments, the most important variable driving treatment heterogeneity is the school average baseline average (aggregated across all three subject tests) with a variable importance between 24 and 70 percent. In particular, there is a sharp difference in treatment effects between schools below and above a baseline average score of 50 percent. In the full sample, other variables of importance include the 2012 Form 2 exam pass rate, baseline student performance and wealth. In the restricted teacher-student matched sample, class size and whether headteachers observe instruction are among the other top 10 most important variables.

research design for both Phase 1 and 2. In addition, we also re-estimate equations (3) and (4) using quantile regressions which allows us to explore both cross-school and cross-student heterogeneity.

In a third part of our empirical analysis we estimate whether the impacts are sustained beyond the intervention period. In order to do this, we re-estimate equations (2) and (4) replacing the dependent variable with the math test scores that were collected from Phase 2 students in late 2015—almost 9 months after they would have been covered by the intervention.

In the last part of our empirical analysis we explore various pathways through which impacts might have been generated. In order to do so, we re-estimate models (1)-(4) replacing the dependent variables with measures of teacher attendance, teacher pedagogical behaviors, and student attendance.

3. Results

3.1 Balance at baseline

Before turning to the analysis of impacts, we first establish that our experimental design is valid by examining the first moments of selected characteristics of schools, teachers, and students in control and treatment groups, using baseline data for phases 1 and 2 (Table 2). We report the p-value of an F-test corresponding to a null hypothesis that there is no difference across the three (in phase 1) and four (in phase 2) arms – clustered at the level of treatment assignment which is the school. In phase 1, the p-values for the test are all, with 1 exception out of the 23 variables, above 0.1 suggesting that the samples are very likely balanced. In phase 2, the p-values for all but 5 variables out of the 23 variables are above 0.1. We nevertheless believe that the samples are well balanced since the absolute magnitudes of the differences across these samples are small in magnitude. For example, the share of schools with electricity is 0.313 in the control group and 0.391 in the arm in which teacher incentives were withdrawn. The largest difference in magnitude is for the variable "headmaster rewards teachers who perform better" where the share is lowest in the student incentives arm (0.533) and highest in the teacher incentives arm (0.677). Importantly, the balance tests suggest that there are no statistically significant differences in the school average baseline test scores of students in the various study arms—either in phase 1 or in phase 2.

3.2 Average impacts on student achievement

Impacts on student learning outcomes in Kiswahili, Math, English, as well as averaged across the three subjects, are reported in Table 3. In phase 1, we find positive and statistically significant learning impacts for the schools that received both teacher and student incentives. These impacts are in the range of 0.14 to 0.18 standard deviation. In phase 2, we find positive and significant learning impacts of teacher incentives which are of a similar—albeit smaller—magnitude (0.09-0.14 sd). We are unable to detect any learning impacts of teacher incentives alone in phase 1 or of student incentives alone in phase 2. Notably, the impacts on learning persist even after teacher incentives are withdrawn. While these are only statistically significant for Kiswahili (column 5) the magnitude of the impacts are similar for the other subjects albeit more imprecisely measured (the effect sizes are all within the range of about 0.1 sd).[21]

It is somewhat unclear why, in phase 1, teacher incentives only have an impact when combined with student incentives, and why, in phase 2, student incentives have no impact when implemented alone. One explanation could be the following. If extra teacher effort is required in order to generate impacts, then that would explain both the lack of impact from student incentives alone as well as the impact from combining them with teacher incentives. But why then did teacher incentives alone not have an impact in phase 1 but did in phase 2? It is possible that during phase 1 the combined intervention created more of an "event" in the school and the program as a whole was therefore more salient to teachers. With time, however, the salience of the teacher incentives, even alone, increased leading to the detectable impacts in phase 2. In addition, it may have taken some time for teachers to view the program as credible, again making it more salient in phase 2.

We take these results as suggesting that low stakes performance-based incentives for teachers have the potential for improving student learning, with modest average impacts. At the same time, student performance-based incentives alone are not likely to lead to appreciable impacts—although combined with teacher incentives they may have promise. Last, we do not find evidence suggesting that the withdrawal of teacher (and student) incentives led to a discouragement effect (in fact we see that impacts seem to persist—albeit with smaller magnitudes).

3.3 Heterogeneity in learning impacts

Table 4 reports the results from estimating models that allow for impacts to differ by a school's average test score at baseline. In phase 1 the results suggest that—while imprecisely measured—teacher incentives alone may have larger impacts on learning outcomes in the initially higher performing schools. The impact in initially poorly performing schools is close to zero and sometimes negative (although statistically insignificantly so), whereas for initially higher performing schools the impact is on the order of magnitude of about 0.04-0.1 sd higher (not significant). The results suggest differentiation in the effect of teacher and student incentives combined, especially for English, where the point estimate on the interaction is a very

[21] Our school-level estimates suggest that two thirds of teachers remain in the school across both phases of the study. This estimate is consistent with considerable teacher turnover following a large secondary school construction program between 2006 and 2010. Across *all teachers,* 22 percent of teachers leave between phase 1 and 2, while new teachers account for just over 38 percent.

substantial 0.45 sd, significant at the 10 percent level. Across all subjects, the average treatment effect differential is 0.2 sd, albeit imprecisely estimated.

The most striking results in Table 4 come from Phase 2. Here we find strong impacts of teacher incentives alone, but these are entirely concentrated in the schools that performed better at baseline. The impacts are statistically significant for all subjects, and the effect sizes are large (0.22 – 0.42 sd). For the average test score across the three subjects, teacher incentives alone led to an increase in learning of 0.33 sd (=0.345-0.019). While not statistically significantly different from 0, the size of the impacts of student incentives alone are also large in magnitude in the initially higher performing schools (with impacts on the order of 0.2 sd) suggesting heterogeneity for this intervention as well. Last, it is for these (initially) high performing schools that withdrawing teacher (and student) incentives seems to have had no negative effect—indeed the effect size for this group is similar to that in which teacher incentives were repeated.

We further investigate heterogeneity by estimating quantile regression models (Tables 5 and 6). The estimates for the size of the impact of teacher and student incentives in phase 1, and of teacher incentives in phase 2, becomes progressively larger as the reference quantile increases. In phase 1 they increase from 0.1 sd at the 10th percentile to 0.21 sd at the 90th percentile (Table 5 Columns 1-5); in phase 2 they increase from 0.02 sd (and statistically insignificant) at the 10th percentile to 0.17 sd (and statistically significant) at the 90th percentile (Table 6 Columns 1-5). This means that students whose (conditional) test scores were lowest exhibited the smallest response to the intervention and that higher performing students were the ones who benefit most.

This inequality increasing effect is magnified when taking into account both cross-school and cross-student heterogeneity (Table 5 and 6 Columns 6-10). As before, teacher incentives only had significant impacts for schools with baseline average test scores above 50 percent —but even among those schools it was students at the higher quantiles for whom impacts are largest. In phase 2, for example, effect sizes increase systematically from 0.152 sd (=0.21-0.058) at the 10th percentile to 0.445 sd (=0.025+0.42) at the 90th percentile. These results are particularly concerning since they suggest that in initially low-performing schools the incentives may have had negative impacts on student learning outcomes. At the 10th percentile of the (conditional) test score distribution, we find that performance-based incentives were associated with statistically significantly lower test scores with effect sizes on the order of -0.06 to -0.07 sd. At other quantiles for these schools, impacts are also generally negative, but smaller and not statistically significantly different from zero.

The impact of the student incentives (either combined with teacher incentives in phase 1, or alone in phase 2) also had heterogeneous impacts across the (conditional) test score distribution—although the patterns are slightly more mixed. As mentioned above, when combined with teacher incentives (phase 1), the size of the impacts increases from 0.1 sd at the 10th percentile to 0.21 sd at the 90th percentile (Table 5 Columns 1-5). This seems to be coming from a combination of impacts at lower percentiles in initially poorly performing schools and impacts at higher percentiles in initially higher performing schools (Table 5 Columns 6-10). In phase 2, the results are much more consistent and suggest that student incentives alone only had an impact in initially higher performing schools—and that this impact did not vary much across the (conditional) test score distribution; impacts are all between 0.21-0.25 sd (Table 6 Columns 6-10).

3.4 Sustainability of learning impacts

The results from the treatment arm in which teacher incentives were withdrawn suggest that program impacts persisted even after the intervention ended. In that particular case it would appear that the additional effort induced by the incentives in phase 1 may have been sustained by the teachers beyond the period for which they were eligible for incentives. In this section we focus on the question of whether students of incentivized teachers, whose learning increased while their teacher was benefitting from the program, were able to sustain that boost to learning as they moved on to a higher grade and another teacher.

Table 7 reports the results for all students (Table 7 Columns 1 and 2) and for the sample restricted to those students who were actually attending the same school in the prior school year and would therefore have been a part of the incentive program if they were in a treated school (Table 7 Columns 3 and 4). The results confirm that the impacts we observed at the end of the year of program exposure were sustained through to the subsequent year. Average impacts of teacher incentives are statistically significant with an effect size of 0.12 sd (when excluding new students). The effects are concentrated among students who were attending schools that performed well at baseline, where the effect size is 0.34 sd (=0.372-0.030). On average we detect no impacts of student incentives, but allowing for heterogeneity reveals a marginally significant negative impact for students in initially low-performing schools (effect size of -0.10 sd) and a positive impact for those in initially high-performing schools (effect size of 0.25 sd (=0.351-0.104)).

3.5 Pathways to learning impacts

The last step in our empirical analysis is to explore pathways through which effects might be generated, focusing on how teachers (and students) adjusted their behaviors to respond to the performance-based incentives. We do this along three dimensions: teacher attendance, teacher classroom behavior, and student attendance. In general, the results do not suggest that incentives operated through these channels.

Consistent with the results from a comparable intervention (Muralidharan and Sundararaman 2011a) we find that teachers did not increase attendance in response to incentives (Table 8). If anything, the results point to negative impacts of teacher incentives on teacher presence, although the sizes of these effects are extremely small (e.g. a reduction of on the order of 4 percentage points compared to a base of about 80 percent). Across domains of teacher in-class behavior, we find significant effects of teacher incentives only in phase 2 in the assignment and review of homework domain. In particular, teachers assigned to the incentive arm have a 0.23 sd larger index on average relative to teachers in the control group (Table 9a Column 8). The point estimate in phase 1 is a modest positive but insignificant effect. Testing for heterogeneity in these effects (Table 9b Column 8) suggests more homework in initially high performing schools, but the point estimates are imprecise. We are unable to detect any statistically significant impacts on the other three classroom behaviors we measure—teacher-student interactions, use of materials, and use of board for lessons and examples (Table 9a and 9b columns 1-3, 5-7). While effect sizes are relatively large, for example on the order of 0.3 to 0.6 for teacher incentive impacts in initially high-performing schools in phase 1 (Table 9b Columns 1-3), these are estimated very imprecisely.

Finally, student attendance does not systematically change as a result of incentives (Table 10). We detect small increases as a response to teacher incentives in phase 1 (on the order of 6 to 8 percentage points compared to a base of 72-75 percent) but these are not consistent with the learning impacts (for example, these impacts are not larger for the initially high-performing schools) and are therefore not a likely pathway for those impacts.

4. Conclusion

This paper examines the impacts of relatively low-cost performance-based incentives to teachers and students in Tanzanian secondary schools. By exploiting a two-phased randomized control trial we estimate average and heterogeneous impacts of incentives. In addition, we examine the effectiveness of incentivizing students (both along with teachers, as well as students only), establish the impact of withdrawing incentives, study the sustainability of those impacts, and explore potential pathways.

We find that the teacher incentives led to modest gains in student learning. Contrary to fears expressed by some policy makers, withdrawal of the performance-based incentives did not lead to discouragement effects. We document that when there were impacts, these were sustained beyond the intervention period. We do not find any evidence that teachers or students responded to incentives by increasing their attendance. Other than increasing homework assignment in phase 2, we find no evidence of other pedagogical behavioral responses across both phases. Finally, incentives aimed only at students did not lead to any average impacts, while those aimed at teachers or teachers and students together were more effective at increasing learning.

We document substantial heterogeneity in impacts which is concerning since the effect is to exacerbate learning inequalities. When there were positive impacts on learning, these were concentrated in schools that were initially high performing. Moreover, we find that impacts were largest for students at the top end of the (conditional) test score distribution. At the same time we find that, in initially lower-performing schools, students, particularly those at the bottom end of the (conditional) test score distribution, may have been negatively impacted by the incentives. So, while performance-based teacher and student incentives may have led to modest average positive impacts on learning, this came at the cost of widening gaps between lower- and higher-performing students and schools and potentially even negatively impacting some students. Note that we cannot identify the mechanism by which inequality is exacerbated. For example, it could be because incentivized teachers concentrate their additional effort on initially higher performing students. Or it could be the case that they provide equal additional effort across all students, but only initially high-performing students benefit from that effort (perhaps because of a curriculum mismatch between the material taught and the distribution of student skills as discussed in Kaffenberger and Pritchett 2020). We cannot distinguish between these mechanisms.

Our findings have both research and policy implications. On the research side they suggest that we need to better understand how curriculum difficulty and the incentives teachers face to prioritize curriculum completion versus student mastery condition teachers' efforts across the student ability distribution. On the policy side, they suggest that future interventions that deploy performance-based incentives—in Tanzania and elsewhere—should take care to mitigate potential inequality-increasing effects.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and student achievement in the Chicago public high schools." *Journal of Labor Economics* 25(1): 95-135.

Atkinson, Adele, Simon Burgess, Bronwyn Croxson, Paul Gregg, Carol Propper, Helen Slater, and Deborah Wilson. 2009. "Evaluating the impact of performance-related pay for teachers in England." *Labour Economics* 16(3): 251-261.

Angrist, Joshua, and Victor Lavy. 2009. "The effects of high stakes high school achievement awards: Evidence from a randomized trial." *The American Economic Review* 99(4): 1384-1414.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and services for college achievement: Evidence from a randomized trial." *American Economic Journal: Applied Economics* 1(1): 136-163.

Athey, Susan, and Stefan Wager. 2019. "Estimating treatment effects with causal forests: An application." *arXiv preprint arXiv*:1902.07409.

Bacache-Beauvallet, Maya. 2006. "How incentives increase inequality." *Labour* 20(2):383–391.

Balch, Ryan, and Matthew G. Springer. 2015. "Performance pay, test scores, and student learning objectives." *Economics of Education Review* 44: 114-125.

Barlevy, Gadi and Neal, Derek. 2012. "Pay for percentile." *American Economic Review* 102(5): 1805–1831.

Barrera-Osorio, Felipe, and Deon Filmer. 2016. "Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches." *Journal of Human Resources* 51(2): 461-499.

Barrera-Osorio, Felipe, and Dhushyanth Raju. 2017. "Teacher performance pay: Experimental evidence from Pakistan." *Journal of Public Economics* 148: 75-91.

Bashir, Sajitha, Marlaine Lockheed, Elizabeth Ninan, and Jee-Peng Tan. 2018. *Facing Forward: Schooling for Learning in Africa*. Africa Development Forum Series. Washington, DC: World Bank.

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin. 2015. "Aligning learning incentives of students and teachers: results from a social experiment in Mexican high schools." *Journal of Political Economy* 123(2): 325-364.

Bettinger, Eric P. 2012. "Paying to learn: The effect of financial incentives on elementary school test scores." *Review of Economics and Statistics 94*(3), 686-698.

Benabou, Roland, and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *The Review of Economic Studies* 70(3): 489-520.

Blimpo, Moussa P. 2014. "Team incentives for education in developing countries A randomized field experiment in Benin." *American Economic Journal: Applied Economics* 6(4): 90-109.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Christophe Rockmore, Brian Stacy, Jakob Svensson, and Waly Wane. 2017a. "What do teachers know and do ? does it matter? Evidence from primary schools in Africa." *World Bank Policy Research Working Paper No 7956.* Washington, D.C. : World Bank

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane. 2017b. "Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa." *Journal of Economic Perspectives* 31(4): 185-204.

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

Chang, Fang, Huan Wang, Yaqiong Qu, Qiang Zheng, Prashant Loyalka, Sean Sylvia, Yaojiang Shi, Sarah-Eve Dill, and Scott Rozelle. 2020. “The impact of pay-for-percentile incentive on low-achieving students in rural China.” *Economics of Education Review* 75:101954.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. “Missing in action: teacher and health worker absence in developing countries.” *Journal of Economic Perspectives* 20(1): 91-116.

Cullen, Julie Berry, and Randall Reback. 2006. “Tinkering toward accolades: School gaming under a performance accountability system.” *Advances in Applied Microeconomics* 14(1): 1-34.

Evans, David K., and Anna Popova. 2016. “What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews.” *The World Bank Research Observer* 31(2):242-270.

Evans, David K., Fei Yuan, and Deon Filmer. 2020. “Are teachers in Africa poorly paid? Evidence from 15 countries.” Unpublished Manuscript.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. “Incentives work: Getting teachers to come to school.” *The American Economic Review* 102(4): 1241-1278.

Figlio, David N. 2005. “Names, expectations and the black-white test score gap.” National Bureau of Economic Research Working Paper No. w11195.

Fryer Jr, Roland. G. 2011. “Financial incentives and student achievement: Evidence from randomized trials” *The Quarterly Journal of Economics 126*(4): 1755-1798.

Fryer, Roland G. 2013. “Teacher incentives and student achievement: Evidence from New York City public schools.” *Journal of Labor Economics 31*(2): 373-407.

Ganimian, Alejandro J., and Richard J. Murnane. 2016. “Improving education in developing countries: Lessons from rigorous impact evaluations.” *Review of Educational Research* 86(3): 719-755.

Gilligan, Daniel, Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M. Lucas, Derek Neal. 2018. “Educator incentives and educational triage in rural primary schools” *National Bureau of Economic Research Working Paper No. 24911*.

Glazerman, Steven, and Allison Seifullah. 2012. “An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years. Final Report.” *Mathematica Policy Research, Inc.*

Glewwe, Paul, and Karthik Muralidharan. 2016. “Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications.” In *Handbook of the Economics of Education*, vol. 5, pp. 653-743. Elsevier.

Glewwe, Paul, Nauman Ilias and Michael Kremer. 2010. “Teacher incentives.” *American Economic Journal: Applied Economics* 2(3):205-27.

Grindle, Merilee. 2004. *Despite the Odds: The Contentious Politics of Education Reform*. Princeton; Oxford. Princeton University Press.

Hanushek, Eric A. 1996. "Measuring investment in education." *The Journal of Economic Perspectives* 10(4): 9-30

Jackson, C. Kirabo. 2010. “A little now for a lot later a look at a Texas advanced placement incentive program” *Journal of Human Resources* 45(3): 591-639.

Jacob, Brian A. 2005. “Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools.” *Journal of Public Economics* 89(5-6): 761-796.

Jacob, Brian A. 2007. “Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments.” *National Bureau of Economic Research Working Paper No. w12817*.

Jacob, Brian A., and Steven D. Levitt. 2003. “Rotten apples: An investigation of the prevalence and predictors of teacher cheating.” *Quarterly Journal of Economics* 118(3): 843-78.

Jennings, Jennifer L., and Andrew A. Beveridge. 2009. “How does test exemption affect schools’ and students’ academic performance?” *Educational Evaluation and Policy Analysis* 31(2): 153-175.

Jinnai, Yusuke. 2016. “To Introduce or Not to Introduce Monetary Bonuses: The Cost of Repealing Teacher Incentives.” *Economics and Management Series*. International University of Japan. http://www.iuj.ac.jp/research/workingpapers/EMS_2016_08.pdf.

Kaffenberger, Michelle and Lant Pritchett. 2020. Failing to Plan? Estimating the Impact of Achieving Schooling Goals on Cohort Learning. RISE Working Paper Series. 20/038. https://doi.org/10.35489/BSG-RISE- WP_2020/038.

Kremer, Michael Esther, Duflo and Pascaline Dupas,. 2007. “Peer effects, pupil-teacher ratios, and teacher incentives: Evidence from a randomized evaluation in Kenya.” Unpublished Manuscript.

Lavy, Victor. 2002. “Evaluating the effect of teachers’ group performance incentives on pupil achievement.” *Journal of Political Economy* 110(6): 1286-1317.

Lavy, Victor. 2009. “Performance pay and teachers' effort, productivity and grading ethics.” *American Economic Review* 99(5): 1979-2011.

Levitt, Steven D., John A. List, and Sally Sadoff. 2016. “The effect of performance-based incentives on educational achievement: Evidence from a randomized experiment.” *National Bureau of Economic Research Working Paper No. 22107*.

Li, Tao, Scott Rozelle, and Linxiu Zhang. 2010. “Cash incentives, peer tutoring, and parental involvement: A study of three educational inputs in a randomized field experiment in China.” *Rural Education Action Project Working Paper No. 221*.

List, John A. , Jeffrey A. Livingston, Susanne Neckermann. 2018. “Do financial incentives crowd out intrinsic motivation to perform on standardized tests?” *Economics of Education Review* 66: 125-136.

Lockheed, Marlaine E., and Adriaan M. Verspoor. 1991. *Improving primary education in developing countries*. Washington, DC. Oxford University Press for The World Bank.

Loyalka, P., Sylvia, S., Liu, C.F., Chu, J., Shi, Y.J. 2019. “Pay by design: Teacher performance pay design and the distribution of student achievement.” *Journal of Labor Economics* 37(3):621-662.

Martin, Gayle and Waly Wane. 2016. “Education Service Delivery in Tanzania.” World Bank Report No. AUS5510. https://microdata.worldbank.org/index.php/catalog/2748/download/39242

Martins, Pedro S. 2009. “Individual teacher incentives, student achievement and grade inflation.” *IZA Discussion Paper No. 4051*.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. “Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania.” *The Quarterly Journal of Economics* 134(3): 1627-1673.

McEwan, Patrick J. 2015. “Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments.” *Review of Educational Research* 85(3): 353-394.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011a. “Teacher performance pay: Experimental evidence from India.” *Journal of Political Economy* 119(1): 39-77.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011b. "Teacher opinions on performance pay: Evidence from India." *Economics of Education Review* 30(3): 394-403.

Neal, Derek. 2008. "Designing incentive systems for schools." *National Center on Performance Incentives Working Paper 2008-16*.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73(2): 417-458.

Springer, Matthew G., John F. Pane, Vi-Nhuan Le, Daniel F. McCaffrey, Susan Freeman Burns, Laura S. Hamilton, and Brian Stecher. 2012. "Team pay for performance: Experimental evidence from the round rock pilot project on team incentives." *Educational Evaluation and Policy Analysis* 34(4): 367-390.

United Republic of Tanzania. 2008. "A Performance Audit Report on School Inspection Programme for Secondary Schools in Tanzania." Dar es Salaam, National Audit Office of Tanzania

Visaria, Sujata, Rajeev Dehejia, Melody M Chao, and Anirban Mukhopadhyay. 2016. "Unintended consequences of rewards for student attendance: Results from a field experiment in Indian classrooms." *Economics of Education Review* 54: 173-84.

World Bank. 2004. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC. The World Bank.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC. The World Bank.

Figure 1: Impact Evaluation Designs

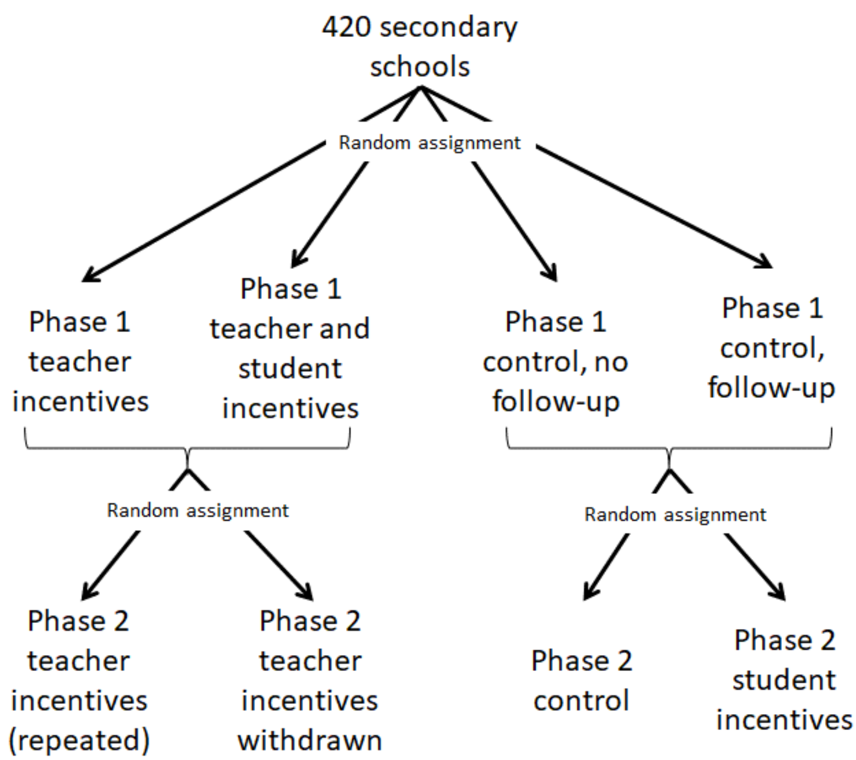

Table 1: Data Collected

Activity	**Timing**	**Number of Schools** Visited	Data Collected
	Phase 1		
Baseline for Phase 1	November 2012-February 2013	420	409
Midline for Phase 1	August – September 2013	304*	304
Endline for Phase 1	November 2013	304	304
	Phase 2		
Baseline for Phase 2	February – March 2014	409	396
Midline for Phase 2	August – September 2014	396	385
Endline for Phase 2	November 2014	396	393
	Post-endline follow-up		
Post-endline test	August-September 2015	396	393

Notes: This table summarizes the timing and number of schools visited and surveyed across three phases of the evaluation. *The overall control group was randomly divided into two groups: one group was not followed up after baseline in Phase 1 (but included in Phase 2), the other was followed-up in each round. The 304 Phase I schools include the latter but not the former

Table 2: Balance Tests

	Phase 1				Phase 2				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Control	TI only	TI+SI	p-value	Control	TI repeated	SI only	TI withdrawn	p-value
Share of schools in Shinyanga	0.561	0.552	0.532	0.918	0.550	0.568	0.529	0.523	0.329
Share Public	0.857	0.792	0.840	0.456	0.885	0.864	0.838	0.769	0.004
Pass Rate	0.511	0.519	0.502	0.972	0.501	0.505	0.488	0.531	0.658
Share with electricity	0.306	0.411	0.355	0.320	0.313	0.328	0.368	0.391	0.040
Share with a generator	0.232	0.221	0.140	0.227	0.213	0.137	0.206	0.234	0.116
Number of toilets for students	11.112	12.281	13.394	0.337	11.219	12.443	11.818	13.435	0.009
Share received capitation grants	0.847	0.792	0.830	0.591	0.870	0.841	0.868	0.800	0.001
Share received other non-capitation grants	0.125	0.146	0.098	0.631	0.179	0.118	0.158	0.127	0.694
Age of the headmaster	39.454	41.660	40.634	0.279	39.843	40.719	39.288	41.557	0.470
Headmaster is female	0.143	0.168	0.106	0.450	0.125	0.148	0.197	0.131	0.788
Headmaster has training in school management	0.531	0.526	0.511	0.961	0.528	0.492	0.431	0.557	0.193
Headmaster rewards teachers who perform better	0.628	0.663	0.681	0.741	0.614	0.661	0.533	0.677	0.005
Female	0.198	0.221	0.212	0.945	0.190	0.241	0.260	0.173	0.231
Years of Experience	10.989	8.917	11.566	0.000	11.792	10.413	13.258	9.826	0.140
Share belongs to Teachers' Union	0.469	0.533	0.451	0.424	0.493	0.479	0.483	0.494	0.998
Share had other source of income last month	0.165	0.177	0.220	0.259	0.211	0.200	0.135	0.186	0.194
Share graduate Teacher	0.286	0.302	0.286	0.913	0.304	0.281	0.340	0.331	0.587
Share more than secondary education	0.843	0.803	0.822	0.513	0.856	0.810	0.818	0.815	0.104
Share Mother can read/write	0.408	0.331	0.400	0.523	0.369	0.371	0.399	0.343	0.920
Share Father can read/write	0.596	0.503	0.579	0.496	0.562	0.556	0.564	0.521	0.959
Average standardized baseline score, Maths	-0.013	-0.021	0.015	0.875	-0.027	-0.086	0.005	0.044	0.334
Average standardized baseline score, English	-0.002	0.012	0.016	0.974	0.003	-0.004	0.038	0.099	0.658
Avg standardized baseline score, Kiswahili	-0.002	0.010	-0.019	0.939	0.007	0.039	0.047	0.047	0.937

Notes: This table presents selected school, teacher and student attributes to illustrate balance of the research design. P-values reported in columns (4) and (9) represent the probability of obtaining the corresponding F-test for a null hypothesis that there is no difference across the three (column 4) and four (column 9) arms. Standard errors are clustered at the school level for this test.

Table 3: Student achievement impacts

	Phase 1				Phase 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Kiswahili	Math	English	Average	Kiswahili	Math	English	Average
Teacher Incentives Only	-0.026	0.056	-0.019	0.002	0.090**	0.144**	0.144*	0.127**
	(0.057)	(0.078)	(0.089)	(0.068)	(0.042)	(0.067)	(0.077)	(0.057)
Teacher and Student Incentives	0.136**	0.158**	0.181**	0.155**				
	(0.058)	(0.080)	(0.091)	(0.068)				
Student Incentives Only					0.039	0.066	0.063	0.057
					(0.051)	(0.071)	(0.083)	(0.062)
Teacher Incentives Withdrawn					0.103**	0.104	0.107	0.105
					(0.044)	(0.081)	(0.095)	(0.067)
Baseline Avg >50	0.445***	0.650***	0.760***	0.621***	0.275***	0.460***	0.532***	0.422***
	(0.051)	(0.080)	(0.086)	(0.066)	(0.043)	(0.070)	(0.076)	(0.059)
N	8296	8299	8318	8529	19171	19208	19182	19239
R-squared	.029	.088	.11	.099	.026	.072	.076	.085

Notes: Table reports estimates from an OLS regression of impacts on *standardized* endline test scores for each phase. Test scores are standardized using the mean and standard deviation of the Control group. Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.

Table 4: Heterogeneity in student achievement impacts: High versus low-baseline schools

	Phase 1				Phase 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Kiswahili	Math	English	Average	Kiswahili	Math	English	Average
Teacher Incentives Only	-0.061	0.016	-0.028	-0.020	-0.003	-0.035	-0.023	-0.019
	(0.072)	(0.065)	(0.086)	(0.065)	(0.044)	(0.051)	(0.064)	(0.046)
Teacher and Student Incentives	0.113*	0.098	-0.015	0.067				
	(0.068)	(0.073)	(0.089)	(0.066)				
Student Incentives Only					-0.012	-0.036	-0.058	-0.034
					(0.056)	(0.050)	(0.069)	(0.051)
Teacher Incentives Withdrawn					0.052	-0.049	-0.047	-0.014
					(0.044)	(0.058)	(0.082)	(0.051)
Baseline Avg >50	0.395***	0.560***	0.578***	0.526***	0.161***	0.218***	0.288***	0.223***
	(0.099)	(0.106)	(0.127)	(0.095)	(0.057)	(0.083)	(0.106)	(0.073)
Teacher Incentives* Baseline Avg >50	0.084	0.102	0.040	0.059	0.220**	0.423***	0.398**	0.345***
	(0.116)	(0.167)	(0.187)	(0.143)	(0.086)	(0.141)	(0.159)	(0.119)
Teach. & Stud. Incent..* Base. Avg >50	0.058	0.145	0.451**	0.203				
	(0.125)	(0.172)	(0.181)	(0.144)				
Student Incentives * Baseline Avg >50					0.122	0.240	0.300	0.221
					(0.117)	(0.179)	(0.199)	(0.151)
Teach. Inc Withdrawn* Base. Avg >50					0.124	0.376**	0.379*	0.294*
					(0.100)	(0.190)	(0.217)	(0.157)
N	8296	8299	8318	8529	19171	19208	19182	19239
R-squared	.068	.16	.209	.202	.043	.112	.129	.142

Notes: Table reports estimates from an OLS regression of impacts on standardized endline test scores for each phase. Test scores are standardized using the mean and standard deviation of the Control group. The dependent variable in columns (4) and (8) is the average standardized score across all three subjects. Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.

Table 5: Heterogeneity in student achievement impacts: High versus low conditional scores and high versus low baseline schools (Phase 1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Teacher Incentives Only	-0.045	-0.038	0.000	0.018	0.051	-0.043	-0.023	-0.021	-0.020	-0.005
	(0.029)	(0.027)	(0.027)	(0.034)	(0.041)	(0.038)	(0.035)	(0.036)	(0.042)	(0.056)
Teacher and Student Incentives	0.100***	0.114***	0.140***	0.171***	0.208***	0.108***	0.093***	0.066*	0.023	0.047
	(0.029)	(0.027)	(0.027)	(0.033)	(0.041)	(0.039)	(0.035)	(0.037)	(0.042)	(0.056)
Baseline Avg >50	0.427***	0.514***	0.610***	0.756***	0.780***	0.448***	0.508***	0.500***	0.577***	0.626***
	(0.026)	(0.024)	(0.024)	(0.029)	(0.036)	(0.046)	(0.042)	(0.044)	(0.051)	(0.067)
Teacher Incentives* Baseline Avg >50						-0.020	-0.044	0.080	0.143**	0.124
						(0.060)	(0.055)	(0.057)	(0.066)	(0.087)
Teach. & Stud. Incent..* Base. Avg >50						-0.041	0.062	0.225***	0.403***	0.321***
						(0.059)	(0.054)	(0.056)	(0.065)	(0.086)
Observations	8529	8529	8529	8529	8529	8529	8529	8529	8529	8529
Pseudo R-squared	0.055	0.069	0.103	0.145	0.156	0.055	0.070	0.105	0.149	0.159

Notes: Table reports estimates impacts of treatments at quantiles of the conditional expected standardized endline test scores for phase 1. The dependent variable is the average standardized score across all three subjects. Test scores are standardized using the mean and standard deviation of the Control group. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment. Baseline Average >50 is an indicator that takes on the value of 1 if the school level mean across all three baseline subject tests is above 50% and 0 otherwise.

Table 6: Heterogeneity in student achievement impacts: High versus low conditional scores and high versus low baseline schools (Phase 2)

	Average impacts					Heterogenous impacts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Teacher Incentives Only	0.020	0.066***	0.108***	0.121***	0.168***	-0.058**	-0.034	-0.012	-0.0032	0.025
	(0.019)	(0.015)	(0.018)	(0.021)	(0.031)	(0.025)	(0.021)	(0.022)	(0.029)	(0.041)
Student Incentives Only	0.009	0.037*	0.057***	0.036	0.009	-0.069**	-0.035	-0.029	-0.031	-0.044
	(0.023)	(0.019)	(0.022)	(0.026)	(0.038)	(0.030)	(0.024)	(0.026)	(0.034)	(0.049)
Teacher Incentives Withdrawn	0.032	0.087***	0.112***	0.096***	0.056	-0.058*	0.0018	0.018	-0.011	-0.043
	(0.022)	(0.019)	(0.021)	(0.026)	(0.037)	(0.030)	(0.025)	(0.027)	(0.034)	(0.049)
Baseline Average >50	0.185***	0.253***	0.383***	0.553***	0.668***	0.054*	0.088***	0.17***	0.34***	0.46***
	(0.017)	(0.014)	(0.016)	(0.019)	(0.028)	(0.028)	(0.023)	(0.025)	(0.032)	(0.046)
Teacher Incent. * Baseline Avg >50						0.21***	0.27***	0.36***	0.36***	0.42***
						(0.039)	(0.032)	(0.034)	(0.044)	(0.063)
SI * Baseline Avg >50						0.22***	0.22***	0.25***	0.21***	0.22***
						(0.049)	(0.041)	(0.044)	(0.056)	(0.081)
TI Withdrawn* Baseline Avg >50						0.23***	0.24***	0.31***	0.33***	0.31***
						(0.047)	(0.039)	(0.042)	(0.054)	(0.078)
N	19239	19239	19239	19239	19239	19239	19239	19239	19239	19239
Pseudo R-squared	0.02	0.034	0.057	0.102	0.134	0.023	0.037	0.061	0.105	0.137

Notes: Table reports estimates impacts of treatments at quantiles of the conditional expected standardized endline test scores for phase 2. The dependent variable is the average standardized score across all three subjects. Test scores are standardized using the mean and standard deviation of the Control group. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment. Baseline Average >50 is an indicator that takes on the value of 1 if the school level mean across all three baseline subject tests is above 50% and 0 otherwise.

Table 7: One year follow up student achievement impacts in Math only (Phase 2 only)

	Including new students		Excluding new students	
	(1)	(2)	(3)	(4)
TI only	0.112**	-0.013	0.122**	-0.030
	(0.056)	(0.043)	(0.059)	(0.046)
SI only	0.043	-0.071	0.030	-0.104*
	(0.067)	(0.052)	(0.070)	(0.055)
TI withdrawn	0.035	-0.086	0.033	-0.084
	(0.079)	(0.056)	(0.080)	(0.056)
Baseline Avg >50	0.229***	0.025	0.230***	0.004
	(0.059)	(0.078)	(0.061)	(0.078)
T1 only* Baseline Avg >50		0.312***		0.372***
		(0.120)		(0.127)
SI only* Baseline Avg >50		0.302*		0.351**
		(0.168)		(0.170)
TI withdrawn* Baseline Avg >50		0.311		0.294
		(0.197)		(0.197)
Constant	-0.226***	-0.133**	-0.224***	-0.110*
	(0.068)	(0.061)	(0.073)	(0.065)
N	29322	29322	25355	25355
R-squared	.035	.0393	.0356	.0411

Notes: Table reports estimates from an OLS regression of the impact on standardized math test scores one year after the Phase 2 endline. Test scores are standardized using the mean and standard deviation of the Control group. Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.

Table 8: Teacher Attendance Impacts in Phases 1 and 2

	Average effects		Heterogeneous effects	
	Phase 1 Teacher is present in school (1)	Phase 2 Teacher is present in school (2)	Phase 1 Teacher is present in school (3)	Phase 2 Teacher is present in School (4)
Teacher Incentives Only	-0.037*	-0.046*	-0.038	-0.017
	(0.022)	(0.025)	(0.029)	(0.030)
Teacher and Student Incentives	0.012		-0.002	
	(0.019)		(0.023)	
SI only		-0.022		-0.011
		(0.025)		(0.031)
TI withdrawn		0.002		0.000
		(0.023)		(0.031)
Baseline Average School Score > 50	0.006	-0.016	-0.009	0.023
	(0.019)	(0.021)	(0.029)	(0.031)
Teacher Incentives* Baseline Avg >50			0.004	-0.084
			(0.044)	(0.057)
Tchr & Stu Incentives* Baseline Avg >50			0.039	
			(0.041)	
SI Only* Baseline Avg >50				-0.035
				(0.053)
TI Withdrawn* Baseline Avg >50				-0.004
				(0.045)
N	3951	6527	3951	6527
R-squared	.0186	.00463	.0192	.0064
Dep Var Mean in Control Group	.747	.737	.747	.737

Notes: Table reports estimates from a linear probability model of the impacts on an unannounced measure of teacher presence at the school. Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.

Table 9a: Classroom Observations – Average Effects

	Phase 1				Phase 2			
	(1) Teacher-student interaction	(2) Use materials	(3) Use of board for lessons and examples	(4) Assign and review homework	(5) Teacher-student interaction	(6) Use materials	(7) Use of board for lessons and examples	(8) Assign and review homework
Teacher Incentives Only	0.022	0.243	0.047	0.198	0.073	-0.029	0.082	0.234**
	(0.139)	(0.165)	(0.141)	(0.154)	(0.098)	(0.101)	(0.097)	(0.094)
Teacher and Student Incentives	0.065	0.105	0.142	-0.163				
	(0.142)	(0.144)	(0.128)	(0.139)				
Student Incentives Only					0.088	0.009	0.180	0.027
					(0.117)	(0.117)	(0.113)	(0.103)
Teacher Incentives Withdrawn					0.039	0.100	-0.150	0.041
					(0.126)	(0.129)	(0.123)	(0.120)
Base Avg School Score > 50	0.229	0.126	-0.240*	-0.131	0.167*	-0.147	0.092	0.119
	(0.140)	(0.171)	(0.141)	(0.142)	(0.101)	(0.094)	(0.088)	(0.086)
N	302	302	302	302	714	714	714	714
R-squared	.176	.0321	.0637	.0426	.0136	.0128	.0228	.0643
Dep Var Mean in the Control Group	-.0245	-.161	.0475	.0380	-.0707	-.0102	-.0201	-.0857

Notes: Table reports estimates from an OLS regression of the impact on the three leading factors drawn from an exploratory factor analysis of the set of classroom observation items used in both phase 1 and 2. The three leading factors are drawn from items related to teacher student interactions (columns 1, 4, 7 and 10); use of textbook materials (columns 2, 5, 8 and 11) and use of the blackboard for lessons, examples and diagrams (columns 3, 6, 9 and 12). Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.

Table 9b: Classroom Observations – Heterogeneous Effects

	Phase 1 (1) Teacher-student inter-action	(2) Use materials	(3) Use of board for lessons and examples	(4) Assign and review homework	Phase 2 (5) Teacher-student inter-action	(6) Use materials	(7) Use of board for lessons and examples	(8) Assign and review homework
Teacher Incentives Only	-0.115	0.047	-0.054	0.201	0.021	-0.092	0.040	0.198*
	(0.164)	(0.166)	(0.177)	(0.189)	(0.116)	(0.117)	(0.121)	(0.113)
Teacher and Student Incentives	-0.016	-0.025	0.128	-0.197				
	(0.157)	(0.159)	(0.137)	(0.169)				
Student Incentives Only					0.128	0.011	0.223	-0.041
					(0.135)	(0.123)	(0.141)	(0.122)
Teacher Incentives Withdrawn					-0.078	0.001	-0.158	0.000
					(0.148)	(0.141)	(0.164)	(0.146)
Baseline Avg School Score > 50	0.021	-0.187	-0.340	-0.166	0.072	-0.268	0.065	0.019
	(0.225)	(0.245)	(0.210)	(0.243)	(0.151)	(0.186)	(0.148)	(0.135)
Tchr Incentives* Baseline Avg >50	0.385	0.550	0.276	-0.002	0.168	0.202	0.128	0.117
	(0.301)	(0.365)	(0.301)	(0.326)	(0.219)	(0.224)	(0.205)	(0.205)
Tchr & Stu Incentives* Baseline Avg >50	0.247	0.401	0.020	0.116				
	(0.339)	(0.364)	(0.325)	(0.307)				
SI Only* Baseline Avg >50					-0.170	-0.027	-0.171	0.246
					(0.267)	(0.293)	(0.225)	(0.227)
TI Withdrawn* Baseline Avg >50					0.326	0.283	0.025	0.128
					(0.269)	(0.296)	(0.246)	(0.261)
N	302	302	302	302	714	714	714	714
R-squared	.19	.0459	.0758	.0433	.024	.02	.0264	.0657
Dep Var Mean in the Control Group	-.0245	-.161	.0475	.0380	-.0707	-.0102	-.0201	-.0857

Notes: Table reports estimates from an OLS regression of the impact on the three leading factors drawn from an exploratory factor analysis of the set of classroom observation items used in both phase 1 and 2. The three leading factors are drawn from items related to teacher student interactions (columns 1, 4, 7 and 10); use of textbook materials (columns 2, 5, 8 and 11) and use of the blackboard for lessons, examples and diagrams (columns 3, 6, 9 and 12). Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.

Table 10: Student Attendance Impacts

	Average effects		Heterogeneous effects	
	(1)	(2)	(3)	(4)
	Phase 1	Phase 2	Phase 1	Phase 2
Teacher Incentives Only	0.052*	0.023	0.085**	-0.011
	(0.027)	(0.025)	(0.035)	(0.031)
Teacher and Student Incentives	0.022		0.028	
	(0.033)		(0.044)	
SI only		-0.013		-0.029
		(0.032)		(0.039)
TI withdrawn		0.005		-0.005
		(0.029)		(0.041)
Baseline Average School Score > 50	0.081***	0.023	0.118***	-0.025
	(0.030)	(0.023)	(0.038)	(0.038)
Teacher Incentives* Baseline Avg >50			-0.083	0.093*
			(0.054)	(0.052)
Teacher & Student Incentives* Baseline Avg >50			-0.020	
			(0.066)	
SI Only* Baseline Avg >50				0.053
				(0.053)
TI Withdrawn* Baseline Avg >50				0.037
				(0.058)
N	10851	18622	10851	18622
R-squared	.022	.0139	.0303	.0162
Dep Var Mean in Control Group	.719	.746	.719	.746

Notes: Table reports estimates from a linear probability model estimating the impact on an unannounced measure of student attendance. Robust standard errors clustered at school level are reported in parentheses. All regressions include a set of region dummies interacted with school quality indicators for the blocks used in random assignment.