Scores, Camera, Action?
Incentivizing Teachers in Remote Areas

Arya Gaduh* — *University of Arkansas*

Menno Pradhan † — *University of Amsterdam, Vrije Universiteit Amsterdam, AIGHD, and Tinbergen Institute*

Jan Priebe ‡ — *GIGA Institute Hamburg, University of Göttingen*

Dewi Susanti § — *World Bank, Jakarta*

January 2019

PRELIMINARY - PLEASE DO NOT CITE OR DISTRIBUTE

Abstract

Poor teacher accountability leads to poor education quality. The problem is especially acute for remotely-located public schools, where school inspectors are less effective in enforcing quality standards. This paper reports the impacts of interventions to improve the accountability of public school teachers in Indonesia's more remote areas. Three interventions combined a pay-for-performance scheme with two sets of tools to monitor and evaluate teacher performance: a *community scorecard* (CSC) and a *tamper-proof camera* to record presence. In all treatments, a community empowerment intervention helped communities develop teacher-specific CSC and form a user committee to evaluate each teacher based on his/her CSC. Treatment 1 had no pay-for-performance component. Treatments 2 and 3 added a pay-for-performance component by linking evaluations to teachers' remote-area allowance. Treatment 2 focuses on absence as the main indicator that determines performance pay and provides schools with a tamper-free camera to record presence. Meanwhile, Treatment 3 evaluates performance based a wide array of indicators (of which presence is one) and does not provide the camera. We find improvements in learning outcomes across all treatments; however, the strongest impacts of between 0.19-0.20 standard deviation (s.d.) were observed for Treatment 2. We find the reduction in teacher absence primarily in Treatment 2. Moreover, we find parental investments in their children's education increased substantially across all treatments.

JEL Classifications: H52, I21, I25, I28, O15

Keywords: Teacher incentives, community-based monitoring, performance pay, remote-area policy

*Sam M. Walton College of Business. Department of Economics. Business Building 402, Fayetteville, AR 72701-1201. Email: agaduh@walton.uark.edu.

†Department of Development Economics, University of Amsterdam and VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: m.p.pradhan@uva.nl.

‡GIGA Institute of Asian Studies, Rothenbaumchaussee 32, 20148 Hamburg, Germany. Email: Jan.Priebe@wiwi.uni-goettingen.de.

§The World Bank, Indonesia Stock Exchange (IDX) Tower 2 L12 & L15, Jalan Jend Sudirman, Senayan, DKI Jakarta 12190, Indonesia. Email: dsusanti@worldbank.org.

1 Introduction

The provision of quality education in public schools in developing countries is constrained by accountability problems. The absence of an established service standard in education, weak parental capacity to demand quality, and poor government supervision all contributed to the problems. These constraints are often much worse in poorer and more remote areas, particularly because top-down supervision from school inspectors is less frequent, and often absent. This leads to a high rate of teacher absenteeism, with negative consequences on student presence, retention, and learning outcomes (ACDP, 2014; Usman et al., 2004; UNICEF, 2012).

In places where top-down supervision is costly, technical solutions and bottom-up, community-based approaches both offer an alternative to monitor service provision. Duflo et al. (2012) find that paying private school teacher in India based on presence as recorded by a tampered proof camera increases absence and student learning. While there is evidence that allowing service users to monitor and provide feedbacks can help improve service quality (Björkman and Svensson, 2009; Barr et al., 2012; Pradhan et al., 2014), the overall evidence on the effectiveness of community monitoring on service delivery improvements have been mixed (Banerjee et al., 2010).

First, all interventions help organize a bottom-up formulation of a joint agreement between teachers, village leadership and parents to improve education quality. Teacher commitments are then formalized through a set of locally defined service standards that include teacher-specific scorecards (or "community scorecards", hereafter, CSC). Second, they also facilitate the formation of a user committee, whose task is to monitor and evaluate each teacher based on his/her CSC on a monthly basis.

The three interventions vary in the tools available to monitor and enforce these service standards. Treatment 1 relies entirely on the aforementioned community empowerment component. Treatments 2 and 3 strengthen the enforcement capacity of the user committee by introducing a pay-for-performance mechanism. In these treatments, CSC performance reports were sent to the district education office, and poor performance led to cuts to teachers' remote-area allowance, a supplemental allowance in the amount of up to one times the teacher's base salary. In Treatment 2 the cut is based on the teacher absence indicator only. It also provided schools with a tamper-proof smartphone camera that teachers use as proof to satisfy the CSC indicator on teacher presence. Meanwhile, Treatment 3 did not prioritize teacher absence in the CSC and linked the allowance cuts to overall scores. It also did not provide schools with the camera.

The study was done in 270 schools in 2 districts in East Nusa Tenggara and 3 districts in West Kalimantan provinces of Indonesia from October 2016 to March 2018. To implement it, we worked closely with the Indonesian Ministry of Education and Culture (MOEC) and the participating district governments. Districts were selected from the central government's list of disadvantaged regions, which were pared down based on a number of cost and implementation considerations. Within each districts, we included schools that satisfy our remoteness criteria and has a minimum number of teachers receiving the remote-area allowance. We then use stratified-random assignment to assign schools to the control and treatment groups.

Our results suggest that these treatments increase learning outcomes. However, while we find that all treatments led to positive improvements in learning — assessed for Indonesian and mathematics

— Treatment 2 shows the most promise. Community empowerment with and without the pay-for-performance component, but without a strong emphasis on absence (Treatments 1 and 3) led to similar improvements in learning. In comparison, the effect sizes in Treatment 2 approximately doubled. Treatment 2 led to improvements in Indonesian and Mathematics by 0.19 and 0.20 standard deviation (s.d.). Overall, these impacts do not differ by gender, but are stronger for students in earlier grades and those who were better performing at baseline.

The treatments increased teacher presence and led to other behavioral changes. Across all treatments, we find that teachers were more likely to be in school when they were supposed to. We also find that teachers were more likely to teach (instead of doing administrative or other work) in class. There is evidence of increased interactions between teachers and parents, as well as between teachers and the school principal. Many of these behavioral changes are driven by teachers who received the remote area allowance.

We also find that treatments lead to significant changes in parental behavior and perception of the school. Treatments increase parental investments in their children's education. Parents in treated communities increased education expenditures and were more likely to find additional support (such as a tutor) for their children. Moreover, parents in treated communities also interacted more with teachers, and were more satisfied with education service delivery and their children's schools in general. These improvements tend to be stronger in Treatment 2 where CSC and the pay-for-performance mechanism are supplanted with the camera. In Treatment 3, however, we find that non-recipients decreased effort in response to the intervention.

The paper contributes to the empirical evidence on how to incentivize teachers in a developing country setting where top-down supervision is costly and often absent. We provide external validity to the idea of linking pay to teacher presence monitored using tamper-proof camera (Duflo et al., 2012) and show it can be implemented in in public schools using allowances paid by the government. As in Cilliers et al. (2018), we find that a monetary incentive tied to teacher presence improves outcomes over local monitoring only.

Our paper also contributes to the design of a performance contract. If outcomes are difficult to measure, it is *a priori* not clear whether the contract should limited to a few objective indicators (such as teacher presence), or a broad measure of performance which is more subjective (Baker et al., 1994). While the former is arguably more fair, it could lead to teachers shifting just focusing on the measured indicators (presence) while reducing effort on other activities that matter for learning (Holmstrom and Milgrom, 1991). The comparison between Treatment 2 and 3 provides an empirical test for the two models and comes out in favor of the former, suggesting that teachers do not trust the user committee to evaluate their performance.

The rest of the paper is organized as follows. The next section discusses the context and the experimental design. Before reporting our results, we briefly discuss how communities respond to the interventions in Section 3. Section 4 discusses the impact of the treatments on student learning outcome, teacher behavior, and parental investments. We discuss both the main impact estimates and the heterogenous impacts of the treatments. Section 5 concludes.

2 Experimental Design

2.1 Teacher Accountability in Remote Indonesia

With the expansion of basic education in Indonesia, the Indonesian government shifted its attention from quantity to quality improvement. So far, a major policy to improve quality was to improve teacher welfare. Half of the national education budget is allocated to pay close to three million teachers' salaries and allowances. Since 2005, the government enacted the Teacher Law that provides an extra "teacher profession allowance" (*Tunjangan Profesi Guru* or TPG) of up to their base salary for teachers who undertake the administrative steps to get themselves certified.[1] Those working in special areas, including remote areas, receive an additional remote-area allowance (*Tunjangan Khusus* or TK) of up to their base salary.[2] However, as none of these allowances are determined by teacher performance, they hardly lead to quality improvement. Studies suggest that recipients of the remote-area allowance are more likely to be absent relative to non-recipients in the same school (SMERU, 2010), and that the professional allowance policy did not improve student learning (de Ree et al., 2018).

Indeed, teacher accountability is a key challenge to improve public education quality, particularly in more remote parts of Indonesia. Consider, for example, the problem of teacher absence. The rate of teacher absenteeism in Indonesia has declined over the past decade, but it remains high in remote areas (19.3 percent) compared to the national rate (9.4 percent) (Usman et al., 2004; ACDP, 2014). High absenteeism rates negatively affects quality, as it increases student absenteeism, drop-out rates, and lowers student learning outcomes (Usman et al., 2004; UNICEF, 2012; Hasan et al., eds, 2013; Suryahadi and Sambodho, 2013). Weak capacity to enforce quality standards, both at the top (government) and the bottom (community) contribute to the lack of improvement.

On the one hand, the government's top-down responses to poor teacher performance has been lacking and in some cases, can further disadvantage more remote schools. Teachers are rarely penalized for their absence from the school. Moreover, for remote schools, logistical challenges to visit implies that they are monitored less often by the school inspectors, who are responsible for addressing a wide range of school-performance issues (UNICEF, 2012). Exacerbating the quality problems for these remote schools, relocation to the remote or poorer areas is often used as a punishment for poorly-performing teachers (Glewwe et al., 2010; Kremer et al., 2005).

On the other hand, bottom-up, community-driven responses in these poor rural communities are limited by the lack of capacity to monitor performance and enforce quality standards. Organizing the oversight of local education services exacts a cost in time and effort, which are in short supply among the poor in rural communities. Furthermore, without information on minimum service standards, education budgets, and the willingness of teachers and principals to act on feedback from communities, communities have little incentive to engage. Barr et al. (2012) suggests that community-driven formulation, monitoring, and enforcement of education service standard indicators can improve learning outcomes.

[1] TPG aims to encourage teacher certification, which requires that teachers to hold an undergraduate degree, submits portfolio of their teaching experiences, and passes a competence test. Once certified, teachers received TPG unconditionally for life, with no recertification necessary. In 2018, close to 1.9 million of elementary and secondary school teachers were TPG recipients, with an annual budget of USD 5.6 billion.

[2] TK covers the additional costs for hardships associated with remote areas. In 2018, close to 69,000 elementary and secondary school teachers were TKG recipients, with an annual budget of USD 183 million.

However, the effects of community monitoring on service delivery improvements have been mixed (see, e.g., Banerjee and Duflo, 2006; Joshi, 2013; Ringold et al., 2012). Moreover, some of these interventions were found to be less effective when later adopted by the government (Duflo et al., 2012; Banerjee and Duflo, 2006; Banerjee et al., 2008).

International evidence suggests that making teachers accountable to their community as opposed to solely being accountable to higher-level education officials can improve service delivery. Several elements are necessary to ensure that community pressure is effective in improving service performance, including (a) having a standard to which the service providers will be accounted for; (b) improving communities' access to information, including their basic rights to services; (c) giving communities the means to influence and voice concerns to service providers; and (d) providing routes to sanction poorly performing service providers (Joshi, 2013; Ringold et al., 2012). In addition, results-based management and performance-based payment could strengthen community demand of quality services (Brinkerhoff and Wetterberg, 2013). We describe next how our design incorporated these elements into the interventions.

2.2 Intervention Design

The Teacher Performance and Accountability interventions (hereafter referred to by its Indonesian abbreviation, *KIAT Guru*) are based on the premises that (1) locally-defined and agreed-upon service standards are more suitable as a benchmark for holding teachers accountable than nationally defined service standards; (2) community members can play important role in monitoring service delivery because they live near the school and have interest in improving service delivery; and (3) community members' inputs may affect teacher efforts differently depending on whether and how they affect the amount of teacher allowance.

This study follows up on an earlier study that tested different ways to strengthen school committees in rural Central Java (Pradhan et al., 2014). This study showed the importance of involving local leadership and ensuring that community involvement leads to concrete actions that improve education. The study also underlined the difficulty of inducing increased efforts of teachers if there are no incentives attached to community action. A pathway analysis suggested that the positive effects on learning in this study were mostly a result of increased inputs of the community and not teacher effort.

The final design for KIAT Guru was informed by an operational pilot conducted in 31 schools in very remote villages in Keerom, Kaimana, and Ketapang districts of Indonesia, from June 2014 to December 2015. The operational pilot tested the implementation of key processes (e.g., facilitation of community meetings, pay-for-performance mechanisms), process-monitoring instruments, and the survey instruments. Key lessons learned from the operational pilot set the parameter for the implementation of the study, particularly on district and village selections.[3]

[3]Among others, we find that the success of the program requires commitments at multiple levels. Community needs to be willing to contribute time and resources and demand better education services. Both district and school managements need to be sufficiently transparent about their finances. Finally, the district bureaucracy needs to be reform-minded enough to fully support program implementation.

2.2.1 Experimental Treatments

There are two core components of our treatments: (i) a community empowerment to formulate service standards and form a user committee to monitor their adherence; and (ii) a pay-for-performance mechanism that links monitoring results to (cuts to) teacher pay. All treatments include the former, but vary in terms of the latter. We first describe each component, followed by the variation that defines the different treatments below.

Community Empowerment Mechanism (CEM). All treatments include a facilitator-driven set of eight meetings. Aside from facilitator visits to the village and the logistical cost for meetings, the pilot did not provide other material or financial investments. The first meeting was a socialization meeting attended by all stakeholders to inform them about the pilot and their rights to participate in education service delivery. It was followed by three separate meetings with representatives of students and alumni, parents and community members, and teachers. These meetings gathered inputs from each stakeholder on how to improve learning environment at school and at home, and what needed to be done by various education stakeholders.

The following meeting brought together all stakeholders to formulate the service agreement and its associated teacher-specific scorecard. The service agreement (SA) is a set of actions to improve the school's learning environments that each stakeholder (i.e., parents, community leaders, teachers, and the school principal) would commit to. From the service agreement, the community scorecard (CSC) is constructed by taking between 5 to 8 indicators that teachers and the school principal committed to. One of the indicators must be teacher presence in school. Based on this shortlist of commitments, the stakeholder meeting would agree on the specific outputs and for each output, assign a weight that reflected (their belief of) its importance to improve learning. The weights must add up to 100. This list, developed for each teacher, became the teacher-specific CSC.

Teacher compliance to the service level defined in the CSC are monitored and evaluated by a User Committee (UC). The UC criteria required a minimum of nine members, with majority of them being female. Three of the members should be community/ religious leaders, and the rest are parents representing each grade level. The facilitation manual was cognizant of other village and school committees and encouraged overlapping memberships. However, as implemented, only a small percentage held memberships in other committees.

In addition to the UC, for each school, a Village Cadre (VC) who is a member of the village community with ideally some experience in community organizing was identified by the facilitator to eventually take over the role of the facilitator. Both the VC and the UC were formally appointed through Village Head decrees, and they were recognized in the district- and national-level regulations as people whose roles were to organize meetings and monitor and evaluate teachers respectively. They received capacity development training from the pilot, and on-the-job mentoring by the facilitators. Their training includes gathering evidence to evaluate teacher service performance in three ways: conducting unannounced visits to the school, interviewing students or teachers, and auditing administrative documents.

After VC and UC members were appointed, a village meeting was conducted monthly to review the implementation of SA and evaluate the CSC. Prior to the meeting, UC conducted a preparation to compile the monitoring results for each teacher. In the meeting, each stakeholder presented their view

about the progress for SA indicators and how efforts could be bettered. Afterward, the UC presented their evaluation of each teacher's CSC for that month, followed with response from the teachers. At the end of the monthly meeting, everyone signed off on the evaluation results, which were then posted or announced in another village meeting and sent to the district government.

After a few months of implementation, a village-wide meeting was held to evaluate the SA, CSC, and the membership of the UC. Prior to this evaluation meeting, the VC and a few of UC members were trained to administer an adaptive Diagnostic Student Learning Assessment (Diagnostic SLA). The Diagnostic SLA aimed to identify students' skills in basic literacy and numeracy along a learning continuum of the national curriculum. After the training, the VC and UC members administer the test to a random sample of six students per grade level. Results from the Diagnostic SLA were shared at the beginning of the evaluation meeting and informed the changes for the SA and CSC indicators.

Pay-for-Performance Mechanism (PPM). Treatments 2 and 3 introduced a pay-for-performance aspect to these interventions. To understand this aspect of the interventions, it is useful to begin with the different types of teachers in our sample schools and their pay components. The majority are public schools, where teacher status are permanent, contract, or honorary teachers. Permanent teachers are tenured civil servants (PNS) whose pay include a base salary and various allowances. An entry level permanent teacher receives around IDR 1.480 million (USD 108), with the most senior one receives around IDR 5.6 million (USD 408) per month. Contract teachers are hired either by district or provincial governments under annual contracts. A contract teacher receives between IDR 1 to 2 million per month (USD 73 to 146). In addition, there are temporary teachers hired directly by the schools, using school operational funds. These temporary teachers have impermanent status, have a much lower pay of between IDR 300,000 (USD 22) to IDR 700,000 (USD 51) per month, and are not eligible for the different types of government teacher allowances. Similarly, in private schools, teachers' employment status can be permanent, contract, or temporary. However, the monthly salary for permanent and contract teachers are lower in private schools compared to public schools.

Our interventions utilized the remote area allowance (TK), which range from IDR 1.5 million (USD 103) and up to one times the teacher's base salary per month. Up to 2016, TK was allocated based on quota from the national government, and proposal from the district government. Starting in 2017, when the pilot just began, TK was allocated based on a national index, with registered teachers appointed to very remote villages automatically receive TK . As such, the pay-for-performance interventions affected only registered teachers, whose employment status can be either permanent or contract, but not honorary. Those who are certified get an additional professional allowance, starting at IDR 1.5 million (USD 103) and up to the teacher's base salary. As such, certified teachers who worked in one of our schools could in theory receive up to three times their base salary.

Our treatments vary how performance evaluations affect the amount of TK allowance received. Eligible teachers in Treatment 1 always received their full TK amount. In contrast, Treatments 2 and 3 differ in the criterion used to link performance with (the cuts to) the TK allowance. In Treatment 2, teacher absence is the prime determinant of the amount of TK received by eligible teachers. Following Duflo et al. (2012), teachers in Treatment 2 schools are provided with a tamper-proof smartphone camera to provide proof of their presence. Teachers take pictures at the beginning and end of a school day and record the times on a manually-entered teacher attendance form.

At the end of each month, members of the UC verify both entries and any letters provided by teachers to account for their absences. There are four types of possible entries, and each determines the total amount cut from their TK. The entry type (daily percent cut) is as follows: full presence (0), partial presence (up to 1.5), excused absence (2), and unexcused absence (5). Once tallied at the end of the month, teachers whose total cut exceeded 15 percent will lose all of their TK for that month.

As such, the CEM for Treatment 2 differed from Treatments 1 and 3 in two ways. First, since a smartphone camera was provided for the school, principal, teachers, school operator, VC, and UC members received training to use it. Second, during the monthly meetings, UC members conducted verification of the camera reports of teacher presence in addition to evaluating teacher CSC.

Meanwhile, in Treatment 3, the scores used to determine the amount of TK received were based on the CSC score. If a teacher received a score of 89 for that month, that she would receive 89 percent of her TK allowance. Even though the CSC is developed and evaluated for all treatment groups, its score only affected the amount of teacher's remote area allowance in Treatment 3 schools.

These TK allowances in all treatment groups were paid on a quarterly basis. TK for civil servant teachers were transferred by the district governments, while TK for the non-civil servants were paid directly by the Ministry of Education and Culture.

A Summary of the Treatments. Table 1 summarizes how our treatments are organized. *Treatment 1* facilitated the development of the service agreement, CSC, and user committee but did not link any of the evaluation results to teachers' TK allowance. As such, teachers in Treatment 1 schools receive the full amount of their TK allowance. *Treatment 2* similarly implemented the community empowerment intervention, but introduced a pay-for-performance scheme where cuts to the TK allowance are only determined by teacher absence. Cameras are used in Treatment 2 to objectively verify teacher presence. Finally, *Treatment 3* implemented a different pay-for-performance scheme: instead of relying solely on teacher presence, cuts to the TK allowance depended on the wide array of indicators listed in the CSC (which would always include teacher presence). Moreover, cameras are not used in Treatment 3.

Table 1: Summary of the Treatments

	Control	Treatment 1	Treatment 2	Treatment 3
CEM: Scorecards and user committee	No	Yes	Yes	Yes
PPM: Presence indicator	No	No	Yes	Yes
PPM: Indicators other than presence	No	No	No	Yes
Tamper-proof camera	No	No	Yes	No

2.2.2 District and School Selection

We work in districts with significant problems of teacher absenteeism in remote, disadvantaged districts. Based on lessons learned from the operational pilot, we exclude districts with very weak governance and with transitory communities (i.e. fishing and the bush communities). To ensure manageable implemen-

tation costs, we excluded districts with very high transportation costs.[4] We also exclude conflict-prone areas, and districts that were part of many other education pilots. Finally, we limit the districts to those that had at least 40 primary schools in rural areas that fulfill our definition of eligible schools described below. Our final list included three districts in West Kalimantan (Ketapang, Sintang, and Landak) and two districts in East Nusa Tenggara (Manggarai Barat and Manggarai Timur).

Schools need to satisfy four eligibility requirements to participate in the study. First, each school must have a minimum of 70 registered students. Second, since the PPM interventions link evaluations to the remote-area allowances, at least 3 of its teachers must receive the remote-area allowance in 2017. Third, schools must satisfy a remoteness criterion of being located in a village that was at least one-hour drive away from the district capital. Finally, we allowed for a maximum of two primary schools (instead of one) per village to be part of the project due to budgetary reasons.[5]

2.2.3 Treatment Assignment

We use a stratified-random assignment procedure to assign schools to control and treatment groups. Each stratum has four villages. The similarity of schools within each stratum is determined by the following variables: village access to a mobile phone signal, the total number of teachers in the school, the share of teachers with the teacher registration number (or "NUPTK"), and the baseline test scores. Villages with two schools were, to the extent possible, grouped with other villages with 2 schools resulting in strata with 8 schools. The last stratum with less than 4 two-school villages was assigned single-school villages instead to complete the assignment. This ensures that two schools in the same village always received the same treatment. Except for this stratum, all other strata had villages with equal number of schools.

We use a simulation to construct groups of similar schools to form a stratum. We begin by constructing a measure of within-group dissimilarity for a particular random grouping of schools. For this, we first standardized all variables by subtracting the mean and dividing by the standard deviation. Then, we define a within-group absolute distance as

$$D(g) = \sum_{k} \sum_{i} \sum_{j,j<i} |x_{gkj} - x_{gki}|$$

where k indexes the underlying matching variable (e.g., the mobile phone signal), i and j denote the village id within the group g. Finally, we sum up the within-group absolute distances across all groups for this random sorting of villages to construct the within-group dissimilarity measure for a particular random grouping.

To determine the groups of schools with the smallest within-group dissimilarity, for each district, we randomly sorted villages, sequentially allocated them to groups, and calculated their total within-group

[4]For example, we exclude Papua, and certain districts in East Nusa Tenggara and Central Sulawesi

[5]To maintain a reasonable implementation budget, we excluded sub-districts (kecamatan) with less than four eligible primary schools and those requiring costly additional travel requirements (e.g. using boat/plane just to reach that specific sub-district). We found less than 270 villages with eligible primary schools. To obtain 270 schools, we needed to have more than 1 school in some of the villages. We therefore randomly chose 170 villages to have a single school participating, and 50 villages to have 2 schools participating in KIAT Guru. In two-school villages, our randomization procedure ensured that both schools received the same treatment. Furthermore, in villages with more than the assigned number of schools, we randomly selected the participating school(s).

dissimilarity. We then take another random draw and repeat this procedure. If the total distance in the new draw is smaller than any in the previous draws, we retain the grouping. We repeated the process 10,000,000 times. Because the procedure is implemented separately for each district, a group is always defined within a district.

2.2.4 Compliance

During the baseline survey, we discovered that three schools in Manggarai Barat were not in the villages suggested by the administrative data used for the initial treatment assignment. In all three cases, these schools were in villages with a school already participating in the study. Since all schools in the same village should be assigned to the same treatment group, we randomly reassigned the treatment status for schools in the three affected villages. The reassignment took place before the start of the intervention.

Moreover, a few weeks before the intervention started, the Ministry of Education and Culture changed the criteria for TK eligibility, from quota-based to a national index for classifying villages, where all teachers located in very remote villages and have been registered becoming eligible. This reclassification excluded three of the villages that were upgraded to remote status. All three schools affected were in the control group. This reclassification implies that the three schools (shown in Table 4) are not "typical" control schools. In our analysis, we introduced a dummy variable to control for these three schools.

2.3 Data Collection

2.3.1 Student Learning Assessments

To evaluate student outcomes, the research team developed its own student learning assessments (SLA) instruments. The instruments assess basic functional literacy (in Indonesian) and numeracy competencies along the learning continuum standards set in the 2006 national curriculum. Designed based on frameworks and findings from other assessment tools (ASER Centre, 2014; Uwezo, 2012; Gove and Wetterberg, 2011; Platas et al., 2014), the developed tools consist of (i) Diagnostic Test which aims to quickly capture students' elementary lower grade (grade 1 to 3) competencies in literacy and numeracy; and (ii) Evaluation Test which maps students' more specific abilities along the literacy and numeracy learning continuum.

We developed a grade-specific test with content drawn from existing test items in the Diagnostic Test and Evaluation Test to avoid having to implement the SLA on a one-to-one basis across all grades. For the baseline survey, SLA was administered for all students in grades 1 to 5 in participating schools, on a one-on-one basis for grades 1 and 2, and on a group basis for grades 3 to 5. At the endline, the SLA was administered to the same set students, the majority of whom were in grades 2 to 6.

2.3.2 Teacher Absence Survey (TAS)

The instrument originated from the World Bank multi-country teacher absence survey (Chaudhury et al., 2006), which calls for an unannounced visit to schools during normal school hours to obtain a representative estimate of teacher absence from school. The instrument has since been adapted for various

TAS implementations in Indonesia. The design and methodology of the KIAT Guru TAS were mainly adapted from Analytical and Capacity Development Partnership (2014) study in Indonesia, with additional input from instrument used in UNICEF (2012) study in Papua and West Papua. The instruments were pre-tested and utilized to gauge the rate of teacher absence from school and classrooms. Information on student absence from school are also collected.

2.3.3 Survey Instruments

In addition to the SLA and the TAS, we collected information from a random sample of respondents from each of the following: (i) school principals, (ii) teachers, (iii) households of children in primary-school-age-attending school, (iv) school committee, (v) the village head, and (vi) the user committee. We collected a rich set of measures to capture their characteristics, perceptions of the education quality and other education stakeholders, as well as the relationships between parents, teachers, school committee members, and the school principal. For parents, we collected detailed information on their monetary and time investments in their children's education. The questionnaires were adapted from previous surveys conducted by the World Bank and others (Hasan et al., eds, 2013; Chu-Chang et al., 2014; World Bank, 2015, 2016; ACDP, 2014).

2.3.4 Data Collection Timeline

An independent survey team collected the baseline and endline survey data, while project facilitators and project implementation team collected the monitoring data. The study started roughly the same time with the school academic year in July 2016. The baseline survey was conducted in October and November 2016 for 213 schools and completed in February 2017 for the remaining 57 schools.

The endline survey was conducted in February until mid-April 2018, soon after the facilitators handed over facilitation to village cadres at the end of 2017. An alternative we considered was to wait with the data collection until Oct 2018. We decided to conduct this round earlier rather than later because we were concerned about fade-outs as a result of the Ramadan, holidays, and class transitions which followed right after (May through June 2018). In addition, we would have lost a cohort of students if we had to wait until after the class transition. The downside of the decision was that the baseline and endline were administered in different months, which could result in seasonality affecting our results.

A qualitative research was also conducted by another group of researchers in nine schools in three districts. They conducted visits to these schools prior to the start of implementation in November 2016, in September 2017 after a few monthly meetings have been implemented, and in March 2018 after project facilitators left.

3 A Note on Implementation

Before discussing our results, we discuss some additional implementation details and report community response to the interventions. These are results from the process monitoring, as part of project management.

3.1 Community Empowerment Mechanism

The set of eight meetings started in November 2016 and completed in June 2017. Process monitoring found that each meeting took about three hours, with the completion of eight meetings ranging between five weeks to three months. Each facilitator was assigned to between four and six schools. In some villages, the initial set of meetings were facilitated by two facilitators due to logistical and personnel safety reasons. The process monitoring and several focus group discussions with facilitators throughout the implementation did not identify differences in how the facilitators conducted these meetings in all treatments.

Service Agreement and the CSC. Since the socialization meeting informed the stakeholders on the rights of children for a safe environment free of physical and verbal abuse, during implementation this indicator became the second most common CSC indicators after teacher presence. Other indicators were on improving learning (teachers conduct various ways to teach and enhance understanding, improve reading, writing and counting, provide additional lessons, provide feedback to students), motivating students, introducing students to social and cultural norms, communicating with parents, and improving teacher behaviors and conducts.

Analysis of the CSC indicators after the evaluation meeting showed some shifts, with an increase in emphasis on improving student literacy and numeracy skills, teachers making lesson plans, and using various learning tools and props. The most significant reduction was on the corporal punishment indicator. Due to geographical challenge and timeline constraint, the evaluation meetings were conducted in 173 schools. For the rest of the schools, the stakeholders were trained to conduct the meetings, and expected to implement them independently. Figure 1 provides the mean CSC scores by treatment. Panel (a) shows the mean scores for all indicators while Panel (b) shows the mean scores based on teacher presence.

User Committee and the Monthly Evaluation Meetings. Most VC personnel and UC members did not change throughout the duration of implementation. About 40 percent of UC members and 30 percent of VC were female, with the majority having a high-school degree. Most village heads also ended up allocating some financial resources to provide a small incentives for the VC and UC members, and overtime pay for the entire monthly meetings .

During implementation, however, variations occurred in how the monthly meeting were conducted. In some villages, discussion between UC members and teachers on the CSC evaluation results were conducted one-on-one. In others, the UC members gave the CSC results to the VC, to be handed over to the teachers. Focus group discussions with the facilitators identified that these differences were influenced by cultural norms, initial resistance from teachers to have their performance being evaluated so openly, and other village-specific sensitivities (unrelated local political economy or social dynamics). By the end of 2017, the facilitators handed over the facilitation of meetings to the VCs.

3.2 Pay-for-performance

During implementation, payment of TK allowances for the pilot schools were more delayed compared to control school, as the governments waited for reports from all schools to come, prior to disbursing.

The payment schedules for non-civil servants were even more delayed, due to higher workload of staff in charge. At the end of 2017, due to the closing of accounting at the end of the calendar and fiscal year, TK for the second half of November and December were paid unconditionally for all treatment groups. A total of 6.7 percent TK for eligible teachers in Treatment 2 was cut, compared to 5 percent in Treatment 3.

4 Results

4.1 Empirical Specification

We estimate the intent to treat (ITT) based on the following regression model:

$$Y_{ijt}^{k} = \alpha_k + \delta Y_{ijt-1}^{k} + \sum_{R} \gamma^r T_j^{kr} + X_{ijt}^{k}\beta + \varepsilon_{ijt}^{k} \tag{1}$$

where Y_{ijt} = the outcome variable for individual i in school j at time $t \in \{0, 1\}$, α_k = the strata fixed effects, and X = control variables. T_j^r is the dummy variable for treatment regime r, and γ^r is the ITT estimate of interest for treatment r. The control variables depend on the outcome of interest, which we elaborate below for each outcome. Our baseline results include estimates with and without the control variables. Standard errors are clustered at the school level. In addition, for robustness, we implement the randomization inference procedure to calculate the *p-values* of the sharp null of no effect for each individual treatment, holding other treatments' assignments constant.

4.2 Summary Statistics and Covariate Balance

Table 2 presents the summary statistics of some key student and school baseline characteristics for the control and treatment groups. One key observation from the table is that the covariates are mostly balanced across control and treatment groups. We find a few statistically-significant differences from the control group for a particular treatment and a particular outcome, which is to be expected from a random assignment. Given generally balanced covariates, we will use the control group statistics to discuss the characteristics of our experimental population below.

Students in our sample have inadequate literacy and numeracy. On average, the mean scores from the SLA's at baseline were 37.8 and 38.6 (out of 100). The student population is quite balanced across gender with 51 percent male. A disproportionate share of students have parents with a primary education. About slightly more than half of the students were taught by teachers who received the remote-area allowance in 2017.

Meanwhile, schools in our sample are quite remote. On average, these schools are located around 40 km from the district education office (UPTD), and it would take an average of 2 hours to reach the education office at a slow speed of 25 km per hour. Teachers live relatively close to their school, living within 3km of the school. Classes in these remote schools tend to be rather small, with a student-teacher ratio of 13.

4.3 Learning Outcomes

Table 3 presents the results for the individual-level regressions of student learning outcomes. Results in odd-numbered columns do not include the control variables. The specification with the control variables includes sex, age dummies, both parents' education, and the school-level mean baseline scores. In addition, we include a set dummy variables for missing baseline outcomes in all specifications, and dummy variables for missing controls in the specification with controls. Finally, we include a dummy variable for whether the village is among the control villages that were reclassified away from the remote status (see Section 2.2.4). Columns 1–4 present the results for the learning outcomes from regressions based on the raw scores, while Columns 5–8 present the results based on the standardized scores.

We find that student learning outcomes improved in all treatments, but the effects were much more pronounced in Treatment 2. We discuss results based on the specification with the control variables (Columns 6 and 8). Treatment 1, where community empowerment and community-based evaluations are not linked to monetary incentives improved Indonesian and mathematics outcomes by 0.10 and 0.07 standard deviation (s.d.) respectively. Linking monetary incentives to the more subjective part of the CSC yielded similar learing impacts of around 0.12 and 0.09 s.d. for Indonesian and mathematics. However, having monetary incentives tied to the objective measures provided by the tamper-proof camera yielded impacts that were around twice as large (0.19 and 0.20 for Indonesian and Mathematics). The *p-values* from the randomization inference procedure are consistent with those from the regressions.

Table 4 presents our heterogeneity analysis by gender, baseline grades, and students' initial SLA scores. To estimate the heterogeneous impacts of the treatments, we estimate the following regression:

$$Y_{ijt}^{k} = \alpha_k + \delta Y_{ijt-1}^{k} + \gamma_h HetVar + \sum_{R} \gamma^r T_j^{kr} + \sum_{R} \gamma_h^r (T_j^{kr} \times HetVar) + X_{ijt}^{k}\beta + \varepsilon_{ijt}^{k} \tag{2}$$

where $HetVar$ is the baseline variable we use for the heterogeneity analysis, and γ_h^r is the differential impact for the subsample of individuals defined by $HetVar$, and the other variables are the same as in Equation 1. Columns 1 and 2 suggest that there is no evidence of heterogenous impacts of the treatments by gender. Columns 3 and 4 show that positive effects of Treatment 2 are more salient for early grade (Grades 1-3) students. Figure 3, which plots the learning-outcome ITT impacts by grade at baseline, supports the conclusion of stronger impacts on lower-grade students. Finally, columns 5 and 6 show that above-median-performing students — to wit, students had better SLA results than their cohort in their class at baseline — benefit more from Treatment 2.

Table 5 suggests that some of these effects may have come from teachers who receive the allowance, but the heterogenous effects are noisily estimated. We examine whether the treatment impacts differ depending on whether in the student's 2016/2017 (baseline) academic year, his/her teacher is an allowance recipient (columns 1–2); whether in his/her 2017/2018 (endline) academic year, his/her teacher is an allowance recipient (columns 3–4); whether the student never had an allowance-receiving teacher in either year; and whether s/he had it in both years. For math, having an allowance-receiving teacher seems to amplify the treatment effects; however, for Indonesian, the results are mixed. In all cases, the coefficients are not statistically significant.

4.4 Teacher Behavior

Table 6 shows results on various on changes in the labor supply of teachers as measured in terms of presence, teaching and working at school. We do not find that teacher presence has increased in any of the treatment groups. However, we observe that the likelihood to work (teach) at school has increased across teachers in Treatment 2 at the 5 (10) percent significance level. Results for Group 1 teachers are similar in magnitude although slightly less significant. Unlike in Treatments 1 and 2, we do not find improvements in labor supply in Treatment 3.

Results from Table 6 mask important differences across teachers receiving and not receiving TK. Tables 7 and 8 report split-sample estimates by teacher TK status. Overall we find that the increases in labor supply in Treatment 1 and in particular in Treatment 2 seem to be exclusively driven by teachers receiving the remote-area allowances.

Finally, we find other behavioral changes among teachers. Table 9 show the impact of the treatments on the number of supervision and parental meetings. We find that teachers are more likely to have met with parents and principals during the last 12 months. While these findings more or less hold across all three treatment groups, they are particular strong for teachers in Treatment 2. We also show in Table 10 that the interventions induce time reallocation among teachers. Across all treatment groups teachers appear to have reduced time on non-teaching tasks, especially the time allocated for pupils homework and exams.

4.5 Parental Investments

Parents in treatment communities invest more money and time in their children's education. Table 11 reports parental investments in their child's education. Education expenditures increased by about Rp 28,000 percent for Treatment 2, while in the control group the average was Rp 324,154. This constitutes an increase of 8.6 percent. For the other treatments the point estimates are smaller, and not significantly different from those observed in the the control or Treatment 2. Across all treatments, parents report that their children receive more support in doing their homework, about 0.35 hours more per week, from a base of about 2.5 hours per week. The fraction of children who had a paid tutor increased by 1.5 percent point only for Treatment 2.

Parents in treatment communities interact more with teachers and are more satisfied with education service delivery. Table 12 reports on the number of meetings between parents and teachers. According parents reports, the number of meetings increased substantially across all treatments. Meetings to discuss learning increased by around 1 to 1.5 per year from a base of 1.5 meetings a year, with the strongest effects observed for treatment 2. For other meetings to discuss other topics, the effects are of similar magnitude, and are the strongest for treatment 3. The results are corroborated by the teacher reports. As discussed above, Table 9 shows that teachers are more likely to have met with parents and principals during the last 12 months. The number of formal meetings teacher have to discuss learning increased by about 0.13-0.25 meetings per month, from a base of 0.36 meetings observed in the control group. Also here, the strongest effects are observed for Treatment 2.

As reported in Table 13, parents are generally satisfied with the school of their child. 91 percent of parents rated the school good or very good. The interventions increased this by about 5 percent point.

For all treatments we find a positive effect of about 21 percent point on the percentage of parents that report that the school is better than last year. The fraction of parents who report that teacher absenteeism or quality is the main problem in education reduced by about 8 percent for Treatments 2 and 3.

The interventions have no significant effect on students attendance, which is already rather high. As reported in Table 14, parents in the control group report their children go to school on average 5.2 days in a week. The observations and attendance records indicate that around 86 to 91 percent of student attend. The estimation point estimates are generally positive, but mostly not significantly different from zero.

5 Conclusion

We present preliminary results from a set of interventions to improve education quality in public schools in remote areas. Our findings suggest that community empowerment, combined with pay-for-performance schemes based on absence (and aided by a tamper-proof camera) and a broad set of performance indicators increased learning outcomes. Interestingly, except for Treatment 2, these interventions did not always adjust teacher behaviors. Instead, we find that these interventions induced behavioral changes among parents: Across all treatments, parents increased their investments in their children's education. In all outcomes, focusing on absence with the aid of the camera yielded the strongest positive results. In future analysis, we would examine further the heterogeneous impacts of the different treatments as well as the some of the mechanisms underlying these findings.

References

Analytical and Capacity Development Partnership, "Study on Teacher Absenteeism in Indonesia 2014," Technical Report, Ministry of Education and Culture 2014.

ASER Centre, "Annual Status of Education Report (Rural) 2013," Technical Report, ASER Centre, New Delhi 2014.

Baker, G., R. Gibbons, and K. J. Murphy, "Subjective Performance Measures in Optimal Incentive Contracts," *The Quarterly Journal of Economics*, November 1994, *109* (4), 1125–1156.

Banerjee, Abhijit and Esther Duflo, "Addressing Absence," *Journal of Economic Perspectives*, February 2006, *20* (1), 117–132.

Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster, "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System," *Journal of the European Economic Association*, April 2008, *6* (2-3), 487–500.

Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy*, February 2010, *2* (1), 1–30.

Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin, "Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda," 2012.

Björkman, Martina and Jakob Svensson, "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," *The Quarterly Journal of Economics*, May 2009, *124* (2), 735 –769.

Brinkerhoff, Derick W and Anna Wetterberg, "Performance-based public management reforms: experience and emerging lessons from service delivery improvement in Indonesia," *International Review of Administrative Sciences*, September 2013, *79* (3), 433–457.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers, "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, February 2006, *20* (1), 91–116.

Chu-Chang, Mae, Sheldon Shaeffer, Samer Al-Samarrai, Andrew B. Ragatz, Joppe De Ree, and Ritchie Stevenson, *Teacher reform in Indonesia: the role of politics and evidence in policy making*, Washington, D.C: World Bank, 2014.

Cilliers, Jacobus, Ibrahim Kasirye, Clare Leaver, Pieter Serneels, and Andrew Zeitlin, "Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools," *Journal of Public Economics*, November 2018, *167*, 69–90.

de Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers, "Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia," *The Quarterly Journal of Economics*, May 2018, *133* (2), 993–1039.

Duflo, Esther, Rema Hanna, and Stephen P Ryan, "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, June 2012, *102* (4), 1241–1278.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, July 2010, *2* (3), 205–227.

Gove, Amber and Anna Wetterberg, "The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy," Technical Report, RTI Press, Research Triangle Park, NC October 2011.

Hasan, Amer, Marilou Hyson, and Mae Chu-Chang, eds, *Early childhood education and development in poor villages of indonesia: strong foundations, later success* Directions in development : human development, Washington, D.C: World Bank, 2013.

Holmstrom, B. and P. Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, January 1991, *7* (special), 24–52.

Joshi, Anuradha, "Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery," *Development Policy Review*, July 2013, *31*, s29–s48.

Kremer, Michael, Nazmul Chaudhury, F. Halsey Rogers, Karthik Muralidharan, and Jeffrey Hammer, "Teacher Absence in India: A Snapshot," *Journal of the European Economic Association*, May 2005, *3* (2-3), 658–667.

Platas, L., L. Ketterlin-Gellar, A. Brombacher, and Y. Sitabkhan, "Early Grade Mathematics Assessment (EGMA) Toolkit," Technical Report, RTI International 2014.

Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, Armida Alisjahbana, and Rima Prama Artha, "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia," *American Economic Journal: Applied Economics*, April 2014, *6* (2), 105–126.

Ringold, Dena, Alaka Holla, Margaret Koziol, and Santosh Srinivasan, *Citizens and Service Delivery: Assessing the Use of Social Accountability Approaches in the Human Development Sectors* Directions in Development, Washington, DC: World Bank, 2012.

SMERU, "Teacher Absenteeism and Remote Area Allowance: Baseline Survey," Technical Report 2010.

Suryahadi, Asep and Prio Sambodho, "Assessment of Policies to Improve Teacher Quality and Reduce Teacher Absenteeism," SMERU Working Paper, SMERU 2013.

UNICEF, ""We Like Being Taught" – A Study on Teacher Absenteeism in Papua and West Papua," Technical Report 2012.

Usman, S., Akhmadi, and Daniel Suryadarma, "When Teachers are Absent: Where Do They Go and What is the Impact on Students?," Technical Report, SMERU 2004.

Uwezo, "Are Our Children Learning? Annual Learning Assessment Report," Technical Report, Twaweza East Africa, Nairobi 2012.

World Bank, "Assessing the Role of the School Operational Grant Program (BOS) in Improving Education Outcomes in Indonesia," Technical Report AUS4133, World Bank, Washington DC 2015.

_ , "Teacher certification and beyond: An empirical evaluation of the teacher certification program and education quality improvements in Indonesia," Technical Report 94019-ID, World Bank, Washington DC 2016.

6 Tables

Table 2: Balance Tables: Individual- and School-Level Variables

	Mean (μ) (standard errors)				Differences = $\mu_{[...]} - \mu_{Control}$ (p-value)		
	Control	Treatment 1	Treatment 2	Treatment 3	Treatment 1	Treatment 2	Treatment 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Individual-Level Variables							
Overall score in Bahasa	37.83	36.94	38.46	36.56	-0.89	0.63	-1.27
	(21.26)	(20.24)	(20.74)	(20.66)	(0.65)	(0.74)	(0.54)
Overall score in Math	38.63	37.14	37.93	36.82	-1.48	-0.69	-1.81
	(22.45)	(21.32)	(21.16)	(21.50)	(0.49)	(0.72)	(0.43)
Male	0.51	0.54	0.52	0.54	0.02	0.01	0.02
	(0.50)	(0.50)	(0.50)	(0.50)	(0.01)**	(0.38)	(0.01)**
Age in years	10.76	10.63	10.69	10.65	-0.13	-0.07	-0.11
	(2.03)	(2.05)	(1.99)	(1.98)	(0.12)	(0.38)	(0.15)
Mother with no education	0.09	0.07	0.11	0.09	-0.02	0.02	-0.00
	(0.29)	(0.25)	(0.32)	(0.29)	(0.19)	(0.51)	(0.93)
Mother with primary education	0.75	0.74	0.71	0.73	-0.01	-0.04	-0.02
	(0.43)	(0.44)	(0.45)	(0.44)	(0.85)	(0.26)	(0.45)
Mother with $>$ primary education	0.15	0.14	0.14	0.21	-0.01	-0.01	0.05
	(0.36)	(0.35)	(0.35)	(0.41)	(0.66)	(0.72)	(0.23)
Father with no education	0.08	0.05	0.09	0.08	-0.03	0.02	0.00
	(0.26)	(0.22)	(0.29)	(0.27)	(0.09)*	(0.48)	(0.96)
Father with primary education	0.71	0.70	0.67	0.69	-0.02	-0.05	-0.03
	(0.45)	(0.46)	(0.47)	(0.46)	(0.59)	(0.13)	(0.30)
Father with $>$ primary education	0.18	0.18	0.18	0.21	-0.00	-0.00	0.03
	(0.39)	(0.38)	(0.39)	(0.41)	(0.94)	(0.97)	(0.41)
TK in 2017	0.53	0.60	0.61	0.62	0.07	0.09	0.10
	(0.50)	(0.49)	(0.49)	(0.48)	(0.14)	(0.10)*	(0.06)*
Use Indonesian to learn at home	0.21	0.24	0.24	0.28	0.04	0.04	0.07
	(0.41)	(0.43)	(0.43)	(0.45)	(0.28)	(0.20)	(0.06)*
Observations	7,011	6,245	6,610	6,714	13,256	13,621	13,725
Panel B. School-Level Variables							
Distance to UPTD (km)	43.48	41.29	44.66	41.56	-2.18	1.18	-1.92
	(35.95)	(36.32)	(39.16)	(39.36)	(0.73)	(0.86)	(0.77)
Time to reach UPTD (minute)	122.99	119.40	124.78	104.07	-3.59	1.79	-18.91
	(114.59)	(111.23)	(98.72)	(63.58)	(0.85)	(0.92)	(0.24)
Travel speed to UPTD (km/hr)	24.70	21.76	22.98	24.63	-2.93	-1.72	-0.06
	(14.87)	(11.59)	(13.28)	(15.95)	(0.20)	(0.48)	(0.98)
Mean teacher travel distance (km)	2.40	3.96	2.16	3.18	1.56	-0.24	0.78
	(3.48)	(9.12)	(4.85)	(4.32)	(0.19)	(0.74)	(0.25)
Mean teacher travel time (hr)	0.26	0.32	0.21	0.25	0.06	-0.05	-0.01
	(0.16)	(0.30)	(0.20)	(0.19)	(0.17)	(0.10)*	(0.63)
Mean teacher travel cost (Rp)	1947.35	2496.66	1842.41	2592.36	549.31	-104.93	645.01
	(2686.68)	(4948.91)	(3359.37)	(3016.55)	(0.43)	(0.84)	(0.19)
Student-teacher ratio (baseline)	13.21	12.25	12.53	12.87	-0.96	-0.68	-0.34
	(4.56)	(4.04)	(3.85)	(5.16)	(0.20)	(0.35)	(0.68)
Excess teacher over learning groups (baseline)	1.88	2.06	1.99	2.34	0.18	0.10	0.46
	(1.54)	(1.93)	(1.69)	(2.69)	(0.55)	(0.71)	(0.22)
Observations	67	68	68	67	135	135	134

Notes: Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 3: ITT on Student Learning Outcomes

	Raw Score				Standardized Score			
	Bahasa Indonesia		Mathematics		Bahasa Indonesia		Mathematics	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment 1	1.297	1.952	1.159	1.520	0.070	0.095	0.055	0.071
	(0.917)	(0.731)***	(0.925)	(0.832)*	(0.044)	(0.035)***	(0.044)	(0.040)*
Treatment 2	3.842	3.949	3.821	4.102	0.185	0.187	0.187	0.201
	(0.987)***	(0.748)***	(1.018)***	(0.856)***	(0.047)***	(0.036)***	(0.049)***	(0.041)***
Treatment 3	1.534	2.444	1.273	1.842	0.086	0.119	0.066	0.094
	(0.868)*	(0.689)***	(0.913)	(0.788)**	(0.041)**	(0.033)***	(0.044)	(0.038)**
Control group mean		47.13		47.03		0.00		0.00
Test of equality (P-val)								
Treatment 1 v. 2	0.013	0.013	0.011	0.004	0.018	0.015	0.008	0.003
Treatment 2 v. 3	0.017	0.048	0.016	0.014	0.028	0.058	0.017	0.014
Treatment 1 v. 3	0.797	0.521	0.905	0.711	0.713	0.504	0.806	0.595
Randomization Inference (P-value, N = 1000)								
Treatment 1	0.193	0.005	0.244	0.081	0.142	0.010	0.202	0.091
Treatment 2	0.002	0.000	0.003	0.000	0.000	0.000	0.002	0.000
Treatment 3	0.119	0.001	0.198	0.028	0.060	0.000	0.149	0.018
R2	0.879	0.888	0.881	0.886	0.290	0.334	0.276	0.301
Observations	31022	31022	31022	31022	31022	31022	31022	31022
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Control variables include sex, age dummies, both parents' education, baseline outcome, dummy variables for individuals with missing baseline outcomes and controls, and the school-level mean scores. The randomization inference tests the sharp null hypothesis of no effect for each individual treatment (holding other treatment assignments constant). Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 4: ITT by Gender, Grades, and Initial Assessment Outcomes

	Male		Grades 4–5		Above-Median Student	
	Indonesian (1)	Math (2)	Indonesian (3)	Math (4)	Indonesian (5)	Math (6)
Subgroup: [...]	-0.220	-0.090	0.113	0.073	0.196	0.149
	(0.024)***	(0.023)***	(0.038)***	(0.042)*	(0.026)***	(0.028)***
Treatment 1	0.080	0.056	0.096	0.099	0.113	0.050
	(0.040)**	(0.044)	(0.045)**	(0.052)*	(0.037)***	(0.045)
Treatment 2	0.186	0.192	0.209	0.239	0.132	0.154
	(0.040)***	(0.045)***	(0.044)***	(0.054)***	(0.040)***	(0.046)***
Treatment 3	0.125	0.074	0.118	0.106	0.086	0.068
	(0.039)***	(0.042)*	(0.044)***	(0.052)**	(0.038)**	(0.042)
Treatment 1 × [...]	0.029	0.030	-0.006	-0.068	-0.035	0.047
	(0.031)	(0.036)	(0.053)	(0.060)	(0.037)	(0.038)
Treatment 2 × [...]	0.002	0.018	-0.111	-0.124	0.066	0.068
	(0.032)	(0.032)	(0.049)**	(0.060)**	(0.036)*	(0.037)*
Treatment 3 × [...]	-0.010	0.038	-0.063	-0.071	0.016	0.020
	(0.033)	(0.035)	(0.051)	(0.060)	(0.038)	(0.038)
Observations	31022	31022	24719	24719	24700	24700
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes
Total subgroup effect:						
Treatment 1	0.109	0.086	0.089	0.031	0.078	0.097
	(0.037)***	(0.044)*	(0.041)**	(0.048)	(0.041)*	(0.045)**
Treatment 2	0.187	0.210	0.098	0.115	0.198	0.222
	(0.039)***	(0.043)***	(0.041)**	(0.048)**	(0.040)***	(0.046)***
Treatment 3	0.114	0.112	0.054	0.035	0.102	0.088
	(0.035)***	(0.041)***	(0.039)	(0.044)	(0.039)***	(0.045)*

Notes: Control variables include sex, age dummies, both parents' education, baseline outcome, dummy variables for individuals with missing controls, school-level mean scores, and a dummy variable for the reclassified villages. Samples are split by whether student mean standardized scores for Indonesian and mathematics at baseline are above/below median. Observations without baseline scores are dropped. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 5: ITT by Teacher TK-Allowance Status

	Baseline Teacher TK17 Recipient		Endline Teacher TK17 Recipient		Never Had TK17 Recipient		Always Had TK17 Recipient	
	Indonesian	Math	Indonesian	Math	Indonesian	Math	Indonesian	Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Subgroup: [...]	-0.020	-0.055	0.041	-0.030	-0.019	0.030	0.041	-0.043
	(0.037)	(0.046)	(0.037)	(0.045)	(0.039)	(0.045)	(0.036)	(0.047)
Treatment 1	0.074	0.055	0.116	0.051	0.092	0.071	0.097	0.054
	(0.046)	(0.057)	(0.045)**	(0.053)	(0.040)**	(0.045)	(0.037)**	(0.045)
Treatment 2	0.135	0.167	0.199	0.159	0.186	0.214	0.198	0.190
	(0.051)***	(0.056)***	(0.048)***	(0.057)***	(0.040)***	(0.045)***	(0.039)***	(0.044)***
Treatment 3	0.081	0.022	0.129	0.070	0.106	0.089	0.124	0.062
	(0.045)*	(0.051)	(0.045)***	(0.053)	(0.035)***	(0.041)**	(0.036)***	(0.041)
Treatment 1 × [...]	0.036	0.031	-0.035	0.035	0.013	0.005	-0.005	0.054
	(0.060)	(0.066)	(0.054)	(0.064)	(0.063)	(0.070)	(0.053)	(0.065)
Treatment 2 × [...]	0.047	0.036	-0.023	0.068	-0.002	-0.053	-0.037	0.035
	(0.056)	(0.066)	(0.056)	(0.067)	(0.066)	(0.073)	(0.048)	(0.066)
Treatment 3 × [...]	0.020	0.095	-0.017	0.038	0.049	0.022	-0.018	0.097
	(0.054)	(0.064)	(0.048)	(0.057)	(0.055)	(0.066)	(0.048)	(0.059)
Observations	24719	24719	31022	31022	31022	31022	31022	31022
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total subgroup effect:								
Treatment 1	0.110	0.086	0.082	0.086	0.105	0.076	0.092	0.108
	(0.044)**	(0.048)*	(0.044)*	(0.050)*	(0.056)*	(0.064)	(0.053)*	(0.060)*
Treatment 2	0.182	0.203	0.177	0.227	0.184	0.161	0.161	0.225
	(0.040)***	(0.051)***	(0.043)***	(0.050)***	(0.060)***	(0.068)**	(0.048)***	(0.064)***
Treatment 3	0.101	0.116	0.113	0.109	0.156	0.111	0.106	0.159
	(0.040)**	(0.049)**	(0.037)***	(0.042)**	(0.055)***	(0.063)*	(0.047)**	(0.056)***

Notes: Control variables include sex, age dummies, both parents' education, baseline outcome, dummy variables for individuals with missing controls, school-level mean scores, and a dummy variable for the reclassified villages. Samples are split by whether student mean standardized scores for Indonesian and mathematics at baseline are above/below median. Observations without baseline scores are dropped. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 6: ITT Impact on Teacher Behavior

	Scheduled at school	Present at school	Working at school	Scheduled to teach	Teaching (wide)	Teaching (strict)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.014	0.021	0.054	0.008	0.050	0.051
	(0.007)**	(0.026)	(0.029)*	(0.021)	(0.031)	(0.031)*
Treatment 2	0.017	0.039	0.058	0.027	0.045	0.053
	(0.006)***	(0.026)	(0.029)**	(0.022)	(0.030)	(0.029)*
Treatment 3	0.011	-0.030	-0.032	0.001	-0.015	0.017
	(0.007)*	(0.029)	(0.032)	(0.024)	(0.034)	(0.031)
Observations	1952	1952	1952	1952	1952	1952
Control Mean	0.9835	0.8202	0.7665	0.7624	0.6116	0.5723
T1vsT2	0.3382	0.4444	0.8873	0.3475	0.8551	0.9214
T1vsT3	0.6042	0.0679	0.0057	0.7374	0.0350	0.2516
T2vsT3	0.1367	0.0109	0.0035	0.2649	0.0488	0.2012
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Baseline controls at school level: # teachers, # pns teachers, # certified teachers,# students, 3 Non-TK schools. Baseline controls at teacher level: Age, gender, marital status, pns and certification status and lagged dependent variable. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 7: ITT Impact on Allowance-Receiving Teachers' Behavior

	Scheduled at school	Present at school	Working at school	Scheduled to teach	Teaching (wide)	Teaching (strict)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.008	0.016	0.052	0.007	0.065	0.064
	(0.005)*	(0.030)	(0.032)*	(0.026)	(0.035)*	(0.034)*
Treatment 2	0.009	0.057	0.097	0.042	0.108	0.111
	(0.004)**	(0.029)*	(0.033)***	(0.027)	(0.037)***	(0.037)***
Treatment 3	0.008	0.014	0.027	0.013	0.059	0.094
	(0.004)*	(0.031)	(0.033)	(0.028)	(0.038)	(0.036)***
Observations	1224	1224	1224	1224	1224	1224
Control Mean	0.9929	0.8286	0.7643	0.7286	0.5536	0.5214
T1vsT2	0.7397	0.1340	0.1579	0.1598	0.1926	0.1690
T1vsT3	0.8764	0.9613	0.4260	0.8246	0.8607	0.3613
T2vsT3	0.5898	0.1273	0.0265	0.2605	0.1657	0.6263
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Baseline controls at school level: # teachers, # pns teachers, # certified teachers,# students, 3 Non-TK schools. Baseline controls at teacher level: Age, gender, marital status, pns and certification status and lagged dependent variable. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 8: ITT Impact on Non-Allowance-Receiving Teachers' Behavior

	Scheduled at school	Present at school	Working at school	Scheduled to teach	Teaching (wide)	Teaching (strict)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.016	0.009	0.047	0.004	0.026	0.034
	(0.015)	(0.044)	(0.047)	(0.037)	(0.051)	(0.049)
Treatment 2	0.033	0.033	0.004	0.021	-0.033	-0.013
	(0.014)**	(0.048)	(0.049)	(0.041)	(0.052)	(0.048)
Treatment 3	0.015	-0.090	-0.127	-0.007	-0.141	-0.108
	(0.016)	(0.051)*	(0.057)**	(0.043)	(0.058)**	(0.056)*
Observations	728	728	728	728	728	728
Control Mean	0.9706	0.8088	0.7696	0.8088	0.6912	0.6422
T1vsT2	0.1290	0.5480	0.3111	0.6710	0.1863	0.2886
T1vsT3	0.9461	0.0188	0.0005	0.7874	0.0010	0.0053
T2vsT3	0.0889	0.0079	0.0145	0.5472	0.0475	0.0681
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Baseline controls at school level: # teachers, # pns teachers, # certified teachers,# students, 3 Non-TK schools. Baseline controls at teacher level: Age, gender, marital status, pns and certification status and lagged dependent variable. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 9: Teacher Meetings with Principals and Parents

	Principal			Parents			
	Monitored	No. meeting	Regular	No. parents	Formal	Informal	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment 1	0.073	0.957	0.077	0.005	0.164	-0.374	-0.174
	(0.037)**	(0.436)**	(0.040)*	(0.084)	(0.075)**	(0.411)	(0.429)
Treatment 2	0.081	0.775	0.079	-0.079	0.246	-0.548	-0.269
	(0.035)**	(0.396)*	(0.043)*	(0.080)	(0.078)***	(0.414)	(0.432)
Treatment 3	0.064	0.744	0.132	-0.141	0.128	-0.476	-0.326
	(0.037)*	(0.422)*	(0.041)***	(0.085)*	(0.069)*	(0.490)	(0.509)
Observations	1593	1593	1593	1593	1593	1593	1593
Control Mean	0.6687	3.8516	0.5894	0.3537	0.3577	2.1850	2.5427
T1vsT2	0.8231	0.6115	0.9704	0.2679	0.3171	0.3437	0.6500
T1vsT3	0.8005	0.5614	0.1313	0.0428	0.6198	0.6655	0.5597
T2vsT3	0.6223	0.9228	0.1720	0.3631	0.1139	0.7600	0.8232
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Baseline controls at school level: # teachers, # pns teachers, # certified teachers,# students, 3 Non-TK schools. Baseline controls at teacher level: Age, gender, marital status, pns and certification status and lagged dependent variable. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 10: ITT impact on teacher time allocation

	Material	Learning	Assess	Training1	Training2	Exams	SelfDev	Scientific	Innovation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment 1	21.977	1.683	-2.250	1.243	-0.777	-1.101	-0.534	-0.155	-0.417
	(13.544)	(1.769)	(0.809)***	(0.590)**	(3.482)	(0.525)**	(0.583)	(0.092)*	(0.315)
Treatment 2	-0.925	3.202	-2.649	0.641	9.229	-1.343	1.145	-0.154	-0.209
	(5.128)	(1.667)*	(0.868)***	(0.547)	(6.416)	(0.545)**	(0.781)	(0.102)	(0.351)
Treatment 3	1.192	0.494	-1.180	0.641	-1.359	-1.400	-0.122	-0.117	-0.158
	(4.598)	(1.690)	(0.945)	(0.598)	(3.931)	(0.584)**	(0.583)	(0.116)	(0.311)
Observations	1593	1593	1593	1593	1593	1593	1593	1593	1593
Control Mean	17.9350	58.5285	14.0650	4.6179	12.9106	6.3537	1.8394	0.3963	0.8496
T1vsT2	0.1056	0.3147	0.5680	0.3001	0.1372	0.6254	0.0321	0.9906	0.2667
T1vsT3	0.1079	0.4394	0.1707	0.3298	0.8880	0.5701	0.4606	0.5216	0.1059
T2vsT3	0.7069	0.0556	0.0823	0.9990	0.1174	0.9190	0.0982	0.5088	0.7855
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Baseline controls at school level: # teachers, # pns teachers, # certified teachers,# students, 3 Non-TK schools. Baseline controls at teacher level: Age, gender, marital status, pns and certification status and lagged dependent variable. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 11: Parental investment in education of children

	Education expenditures		Hours support homework		Child has paid tutor	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	16,117	14,185	0.336	0.321	-0.001	-0.001
	(13,808)	(13,530)	(0.188)*	(0.188)*	(0.003)	(0.003)
Treatment 2	28,658	27,495	0.378	0.359	0.015	0.015
	(14,162)**	(13,823)**	(0.189)**	(0.189)*	(0.006)***	(0.006)***
Treatment 3	10,457	8,274	0.349	0.333	0.004	0.004
	(14,337)	(13,981)	(0.196)*	(0.193)*	(0.004)	(0.004)
Control group mean	324,154		2.46		0.00	
Test of equality (P-val)						
Treatment 1 v. 2	0.349	0.319	0.804	0.824	0.006	0.006
Treatment 2 v. 3	0.170	0.571	0.874	0.882	0.058	0.063
Treatment 1 v. 3	0.663	0.986	0.938	0.945	0.253	0.231
R2	0.727	0.732	0.416	0.428	0.057	0.060
Observations	5386	5386	5394	5394	5401	5401
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Notes: Education expenditures are for the past school year. Hours support homework is the number of hours the child received support for education work from someone in the household in the past week. Three control schools were no longer TK-eligible during the intervention period. All regressions control for the baseline value of the dependent variable, in interaction with a dummy variable indicating if the baseline value is missing. Controls include sex of the child, age of the child, education level of both parents, a dummy indicating whether the mother is the respondent and the age of the respondent. It also includes dummy variables for each control variable indicating whether it is missing. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 12: Number of meetings with the principal or a teacher in the last academic year to discuss

	Learning outcomes		Other	
	(1)	(2)	(3)	(4)
Treatment 1	1.264	0.742	1.419	0.775
	(0.259)***	(0.265)***	(0.263)***	(0.256)***
Treatment 2	1.685	1.103	1.715	0.946
	(0.302)***	(0.287)***	(0.280)***	(0.270)***
Treatment 3	1.448	0.929	1.845	1.156
	(0.336)***	(0.338)***	(0.328)***	(0.326)***
Control group mean	1.50		0.73	
Test of equality (P-val)				
Treatment 1 v. 2	0.165	0.232	0.387	0.616
Treatment 2 v. 3	0.517	0.632	0.736	0.582
Treatment 1 v. 3	0.578	0.573	0.251	0.293
R2	0.203	0.251	0.124	0.202
Observations	5401	5401	5401	5401
Strata FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: Three control schools were no longer TK-eligible during the intervention period. All regressions control for the baseline value of the dependent variable, in interaction with a dummy variable indicating if the baseline value is missing. Controls include a dummy for membership in the user committee, sex of the child, age of the child, education level of both parents, a dummy indicating whether the mother is the respondent and the age of the respondent. It also includes dummy variables for each control variable indicating whether it is missing. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table 13: Parent satisfaction with education service delivery

	Good or very good		Better than last year		Teachers Main Problem	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.044	0.045	0.200	0.202	-0.049	-0.054
	(0.018)**	(0.018)**	(0.028)***	(0.028)***	(0.031)	(0.031)*
Treatment 2	0.049	0.048	0.225	0.224	-0.072	-0.073
	(0.017)***	(0.017)***	(0.027)***	(0.027)***	(0.031)**	(0.031)**
Treatment 3	0.047	0.048	0.200	0.200	-0.080	-0.082
	(0.018)***	(0.018)***	(0.028)***	(0.028)***	(0.031)**	(0.031)***
Control group mean	0.91		0.44		0.27	
Test of equality (P-val)						
Treatment 1 v. 2	0.691	0.785	0.354	0.392	0.373	0.468
Treatment 2 v. 3	0.904	0.970	0.347	0.344	0.776	0.730
Treatment 1 v. 3	0.783	0.816	0.997	0.946	0.225	0.271
R2	0.950	0.950	0.625	0.627	0.277	0.287
Observations	5285	5285	5276	5276	5401	5401
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Notes: Three control schools were no longer TK-eligible during the intervention period. All regressions control for the baseline value of the dependent variable, in interaction with a dummy variable indicating if the baseline value is missing. Controls include sex of the child, age of the child, education level of both parents, a dummy indicating whether the mother is the respondent and the age of the respondent. It also includes dummy variables for each control variable indicating whether it is missing. Standard errors clustered at the school level. */**/*** denotes significant at the 10/5/1 percent significance levels.

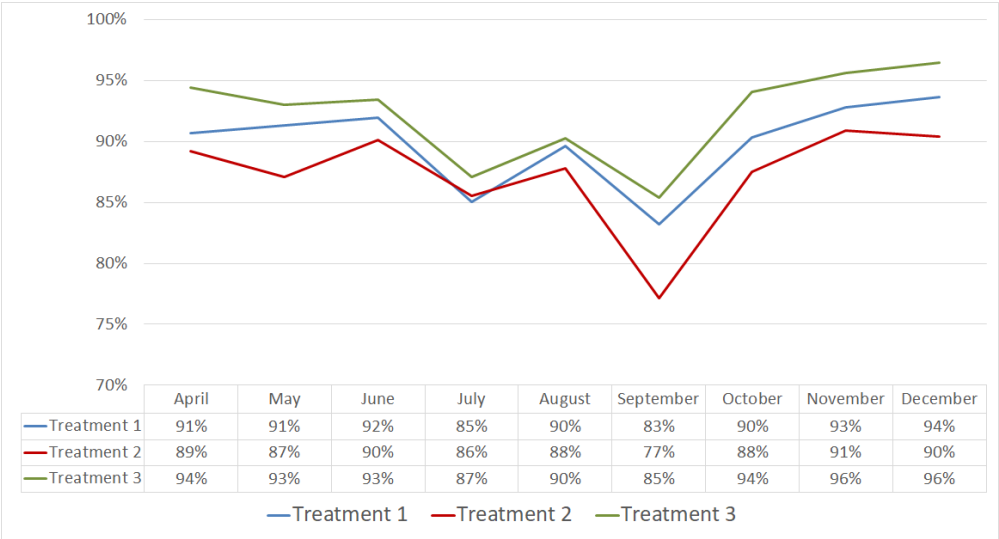
Table 14: Student attendance

	Nr of school days in past week		Fraction of students in school (observed)		Fraction of students in school (recorded)	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 1	0.155	0.160	0.033	0.039	0.016	0.015
	(0.122)	(0.120)	(0.018)*	(0.018)**	(0.011)	(0.011)
Treatment 2	0.199	0.203	0.023	0.027	0.022	0.020
	(0.127)	(0.125)	(0.017)	(0.018)	(0.011)**	(0.011)*
Treatment 3	0.190	0.196	0.025	0.023	0.010	0.009
	(0.120)	(0.119)	(0.018)	(0.018)	(0.011)	(0.011)
Control group mean	5.20		0.86		0.91	
Test of equality (P-val)						
Treatment 1 v. 2	0.697	0.704	0.570	0.498	0.571	0.607
Treatment 2 v. 3	0.962	0.951	0.908	0.826	0.250	0.317
Treatment 1 v. 3	0.702	0.723	0.656	0.379	0.553	0.619
R2	0.960	0.961	0.990	0.990	0.996	0.997
Observations	5392	5392	270	270	270	270
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Notes: Number of school days in past week as reported by parents. Observed number of students present is collected by interviewers through classroom observation. The fraction is calculated by dividing by the number of students observed in school by the number of students enrolled. In the same section, the school official also fills in the number of students that did not attend that day. The recorded number of students is constructed by subtracting the absentees from the number of enrolled students. Three control schools were no longer TK-eligible during the intervention period. All regressions control for the baseline value of the dependent variable, in interaction with a dummy variable indicating if the baseline value is missing. Controls in column 2 include sex of the child, age of the child, education level of both parents, a dummy indicating whether the mother is the respondent and the age of the respondent. Controls in column 4 and 6 include total number of students, fraction of PNS teachers and distance from school to Subdistrict Education UPTD (KM). The regressions also include dummy variables for each control variable indicating whether it is missing. */**/*** denotes significant at the 10/5/1 percent significance levels.

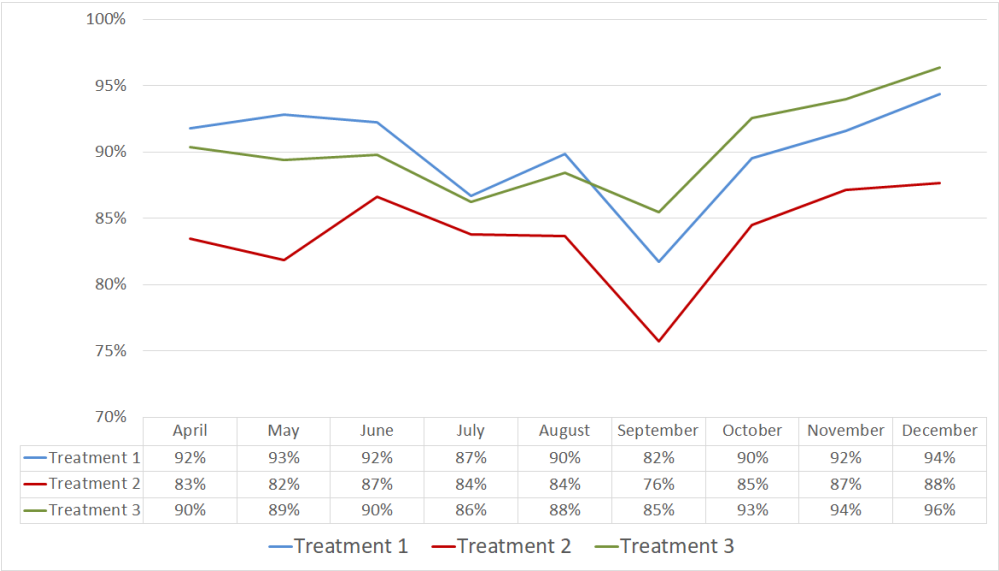
7 Figures

Figure 1: The Evolution of the CSC Scores



	April	May	June	July	August	September	October	November	December
Treatment 1	91%	91%	92%	85%	90%	83%	90%	93%	94%
Treatment 2	89%	87%	90%	86%	88%	77%	88%	91%	90%
Treatment 3	94%	93%	93%	87%	90%	85%	94%	96%	96%

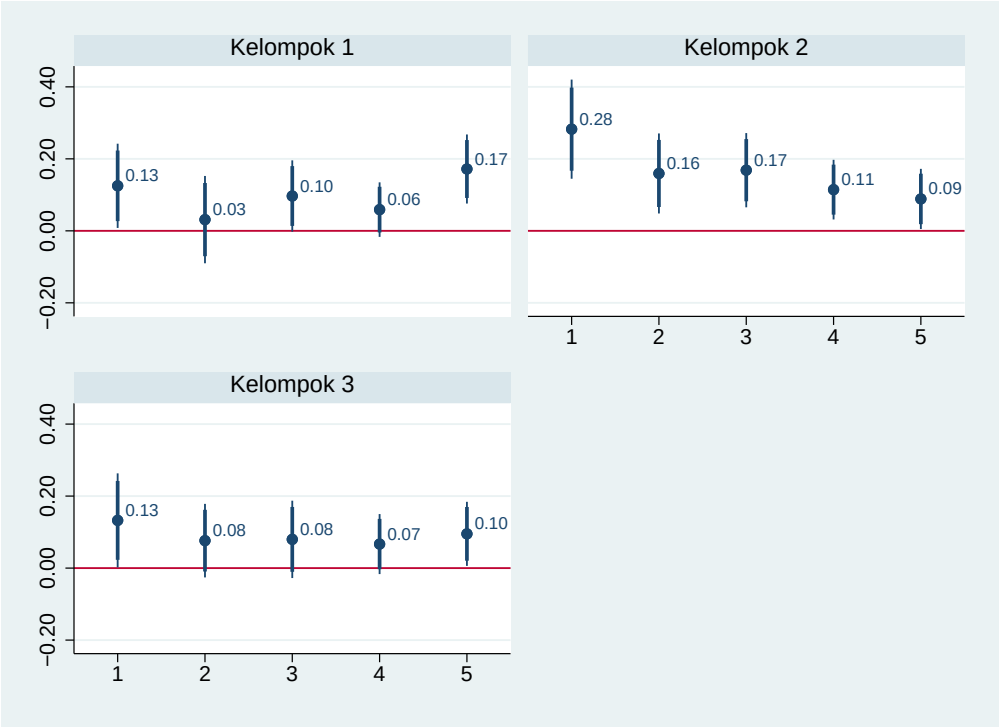
(a) Overall Indicator



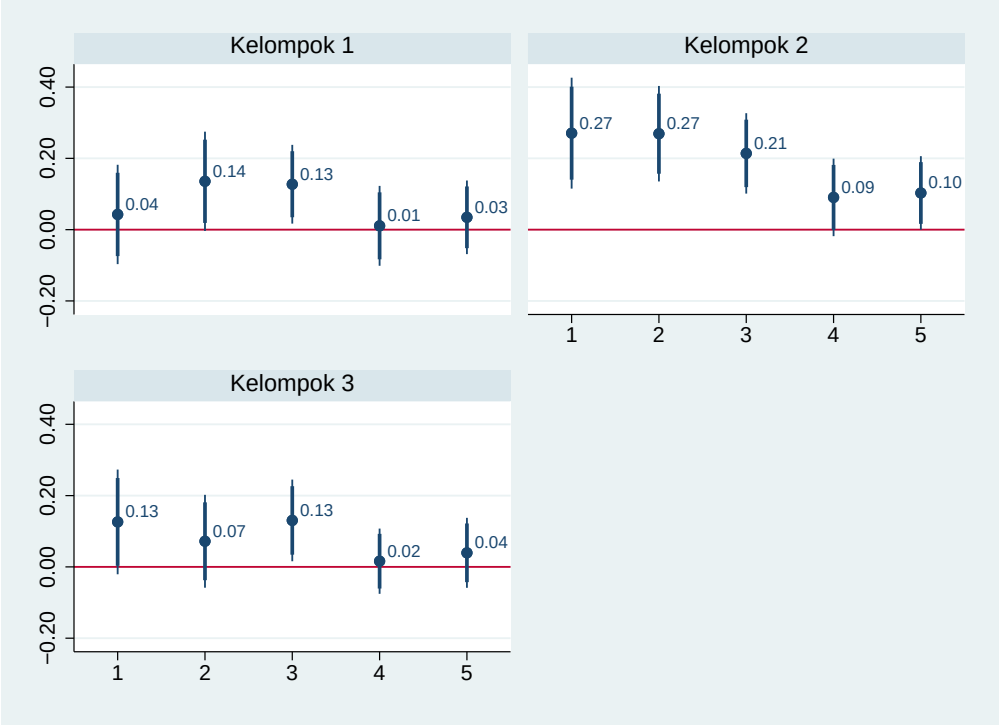
	April	May	June	July	August	September	October	November	December
Treatment 1	92%	93%	92%	87%	90%	82%	90%	92%	94%
Treatment 2	83%	82%	87%	84%	84%	76%	85%	87%	88%
Treatment 3	90%	89%	90%	86%	88%	85%	93%	94%	96%

(b) Teacher Presence

Figure 3: ITT on Standardized SLA Score by Baseline Grade Level



(a) Indonesian



(b) Mathematics

Appendix

A Community Empowerment

The CEM provides community members with an explicit role to monitor and evaluate teacher service performance and to ensure teacher accountability. The CEM is implemented by project facilitators, each is responsible for five to six villages. They initially conducted the meetings, but later built the capacity of Kader Desa (village cadres) to facilitate meetings and provided them with on-the-job mentoring. The CEM finances: (i) project facilitators; (ii) training and coaching of Kader Desa; (iii) capacity building for parents and community members through trainings and coaching by project facilitators and Kader Desa; (iv) project-facilitator-facilitated meetings between community members and teachers to develop Service Agreements (SA) and Community Score Card (CSC); (v) meeting for community members in selecting User Committee (UC); (vi) monthly meetings to discuss UC's evaluation of teacher service quality and community members' evaluations on education service improvements; and (vii) transmitting the UC's evaluation and consolidated inputs to district education authorities.

The SA and CSC were developed through the following sequence of meetings led by project facilitators and Kader Desa: (i) with upper grade students and alumni representatives to discuss their wishes for ways to improve learning environments at school and at home; (ii) with parents and community representatives to present their children's wishes, and discuss what should be done by teachers, parents, and community representatives to improve learning environments at school and at home; (iii) with principal and teachers to cover the same topics; (iv) with principal, teachers, parents, and community members to agree upon a list of each of their promises to improve learning environments at school and at home; the SA listed the promises from each stakeholder; the CSC is shortlisted from principal's and teachers' promises in the SA, with specific outputs and weights. Principal, teachers, parents, and community representatives were mostly free to choose the indicators to be included in the SA and CSC. However, facilitators were instructed to encourage indicators that relate to student learning and that are quantifiable. In addition, indicator on teacher presence must be included in all SA and CSC.