

Double for Nothing?

Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia

Joppe de Ree

Karthik Muralidharan

Menno Pradhan

Halsey Rogers



WORLD BANK GROUP

Education Global Practice Group

December 2017

Abstract

How does a large unconditional increase in salary affect the performance of incumbent employees in the public sector? This paper presents experimental evidence on this question in the context of a policy change in Indonesia that led to a permanent doubling of teachers' base salaries. The analysis uses a large-scale, randomized experiment across a representative sample of Indonesian schools that accelerated this pay increase for teachers in treated schools. The findings show that the large pay increase significantly improved teachers' satisfaction with their

income, reduced the incidence of teachers holding outside jobs, and reduced self-reported financial stress. Nevertheless, after two and three years, the increase in pay led to no improvement in student learning outcomes. The effects are precisely estimated, making it possible to rule out even modest positive impacts on test scores. The results suggest that unconditional pay increases are unlikely to be an effective policy option for improving the effort and productivity of incumbent employees in public sector settings.

This paper is a product of the Education Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at joppederee@gmail.com, kamurali@ucsd.edu, m.p.pradhan@vu.nl, and hrogers@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia[†]

JOPPE DE REE

KARTHIK MURALIDHARAN

MENNO PRADHAN

HALSEY ROGERS

JEL Codes: J31, J45, I21, C93, O15.

Keywords: Indonesia, Education, Public sector management, Public service delivery, Teacher policy, Teacher compensation, Labor markets, Randomized trials

[†] We are especially grateful to Gordon Dahl and Lawrence Katz (the editor) for extensive comments on multiple drafts of this paper. We also thank Nageeb Ali, Eli Berman, Julie Cullen, Uri Gneezy, Roger Gordon, Gordon Hanson, Richard Murphy, Derek Neal, Ben Olken, Hessel Oosterbeek, Valerie Ramey, Rivandra Royono, Ritchie Stevenson, Miguel Urquiola, and several seminar participants for comments. We are grateful to the Indonesian Ministry of Education and Culture for its interest in evaluating its teacher pay reforms, and for supporting this large-scale experiment and data collection. This evaluation would not have been possible without generous financial support from the government of the Kingdom of the Netherlands. The authors are grateful to Dedy Junaedi (and team), Titie Hadiyati (and team), Susiana Iskandar, Amanda Beatty, and Andy Ragatz for their exceptional efforts and support in conducting this evaluation as part of the World Bank BERMUTU project team at various points of time over the course of this project, and to counterparts at the Indonesian Ministry of Education and Culture, including Dr. Baedhowi, Dian Wahyuni, Santi Ambarukmi, Yendri Wirda Burhan, Simon Sili Sabon (and the team at *pustlitjak*), Dhani Nugaan, Bastari, Hari Setiadi, Rahmawati, and Yani Sumarno (and the team at *puspendik*), who supported this experiment and implemented it flawlessly. Over the years, the project also benefited from excellent research assistance of Ai Li Ang, Husnul Rizal, and others at the World Bank office in Jakarta. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the National Bureau of Economic Research, the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

I. INTRODUCTION

The level and structure of public-sector compensation play a key role in the ability of governments to attract, retain, and motivate high-quality employees, and to deliver services effectively (Finan, Olken, and Pande 2017). As a result, countries sometimes implement large increases in public-sector salaries to attract higher-quality applicants to government jobs and to better motivate existing employees (see Govt. of India 2008 and 2015 for instance). While such salary increases may improve the quality of new employees hired over time, they also lead to substantially higher salary spending on existing employees, with large fiscal costs that crowd out other public expenditure.¹ Thus, understanding the extent to which unconditional pay increases make incumbent public-sector workers more motivated and productive is a key consideration in evaluating the cost effectiveness of such salary increases.

Yet there is limited evidence on this policy-relevant question, in part because conducting empirical research in public-sector personnel economics is difficult. Challenges include measuring employee productivity in the public sector, and generating exogenous variation in the pay of public-sector workers. A growing experimental literature examines how changes in public-sector compensation affect worker productivity, but most studies to date have focused on pilots of performance-linked bonus programs, as opposed to the unconditional pay increases that are much more typical in bureaucracies (see Finan, Olken, and Pande 2017 for a review).

In this paper, we provide experimental evidence on the impact of a large unconditional salary increase on the effort and productivity of incumbent public employees. Our study was conducted in the context of a policy change in Indonesia that permanently doubled the base pay of eligible civil-service teachers who went through a certification process.² The reform moved teacher salaries from the 50th to the 90th percentile of the college-graduate salary distribution. Civil-service teachers in

¹ Compensation to government employees constitutes one of the largest items of public expenditure in most countries, representing an average of 24.5% of total government spending in high-income countries, and 27% in low and middle-income countries (International Monetary Fund 2016). In labor-intensive sectors like health and education, the average share of salaries in government spending rises to 42.8% and 65.6% respectively (Clements, Gupta, and Nozaki 2013). Thus, across-the-board salary increases are very expensive. For instance, the unconditional salary increases to government workers awarded by the recent 7th Pay Commission in India raised government expenditure by 0.65% of GDP and required foregoing or deferring capital investments to meet fiscal deficit targets (Govt. of India 2015; Sabnavis and Shah 2015).

² The policy was designed to reward a process of teacher skill upgrading (signaled by "certification") by providing a certification allowance that was equal to the base pay (thereby doubling base pay). However, in practice, the certification mainly consisted of the pay increase (see section II for details).

Indonesia also enjoy generous benefits and high job security, and quit rates were very low even before the pay increase. Thus, the teachers in our study are typical of public-sector employees in many low- and middle-income countries, who hold highly coveted jobs and enjoy a significant wage premium relative to their private-sector counterparts with similar observable characteristics (Finan, Olken, and Pande 2017).

Given the large fiscal burden of the policy, teacher access to the certification program was phased in over 10 years (from 2006 to 2015), with priority in the queue being determined by seniority. Thus, many "eligible" teachers had to wait several years before being allowed to enter the certification process. Working closely with the Government of Indonesia, we implemented an experimental design that took advantage of this phase-in. It allowed all eligible teachers in 120 randomly selected public schools to access the certification process and the resulting doubling of pay *immediately*; in contrast, teachers in control schools experienced the "business as usual" access to the certification process through the gradual phase-in over time. The study was conducted over a three-year period, in a near-nationally representative sample of 360 schools drawn from 20 districts and all major regions of Indonesia.

Our experiment successfully accelerated access to the certification process and doubling of pay for eligible teachers in treatment schools. It resulted in a 29-percentage-point increase in the fraction of teachers in treatment schools who had been certified and received the salary supplement at the end of two years, and a 24-percentage-point increase at the end of three years (relative to the control group).³ Among the "target" teachers (who were eligible but not certified at the baseline), there was a 54 (and 45) percentage-point increase in teachers who were certified and received their salary supplement at the end of two (and three) years in treatment schools (relative to control schools).

The experiment significantly improved measures of teacher welfare: At the end of two and three years of the experiment, teachers in treated schools had higher income, were more likely to be satisfied with their income, and were less likely to report financial stress. They were also less likely

³ Roughly 20% of teachers in both treatment schools were already certified at baseline, and another 25% of teachers were not eligible for certification at baseline, because they were either not civil-service teachers or not college graduates. It is the remaining 55% of teachers who were "eligible but not certified" at the baseline (whom we describe as "target" teachers) who were affected by the experiment, and it is this population of teachers for whom the experiment accelerated access to certification and induced a significant increase in pay. Note that the "first stage" of the experiment weakens over time as the certification rate in the control group catches up with that in the treatment group.

to hold a second job, and they worked fewer hours on second jobs (the last two differences are significant after two years, but not after three).

Yet, despite this improvement in incumbent teachers' pay, satisfaction, and time available to focus on their main job (due to a reduction in second jobs), the policy did not improve either their effort or student learning. Teachers in treated schools did not score better on tests of teacher subject knowledge, and we find no consistent pattern of impact on self-reported measures of teacher attendance. Most importantly, we find no difference in student test scores in language, mathematics, or science across treatment and control schools. The point estimates are close to zero and precisely estimated, allowing us to rule out effects as small as 0.05 standard deviations (σ) at the 95% level in treated schools. We present non-parametric plots of quantile treatment effects and find no effect on test scores in treated schools at any point in the test-score distribution. Finally, we use the school-level random assignment as an instrumental variable for being taught by a certified teacher in a given year, and find no improvement in student test scores from being taught by a certified teacher (relative to students in control schools taught by similar "target" teachers). These effects are also precisely estimated, allowing us to rule out effects larger than 0.1σ at the 95% level.

Our results suggest that several posited mechanisms by which an unconditional salary increase could lead to improved effort and productivity of incumbent workers may not have applied in our setting. For instance, it is often argued that increasing employee pay in non-incentivized pro-social tasks like teaching or health care may reduce time spent on outside jobs and increase time and effort on the primary job (UNESCO 2014). Advocates of higher pay also point to models of reciprocity and gift-exchange where employees pay back employers for a wage premium with an effort premium (Akerlof 1982). Finally, qualitative studies have argued that low pay makes it difficult for managers to demand accountability from employees who are considered underpaid, and that higher pay would foster greater professionalism and adherence to standards (Webb and Valencia 2006).

It is important to note that our results are from a large-scale experimental evaluation of a *policy* change that aimed to improve education quality. By design, such policy experiments are unlikely to yield a precise theoretical test of any one of the mechanisms listed above.⁴ However, from a policy

⁴ For instance, reciprocity may require that the "gift" of a higher salary be received from an employer whom the employee interacts with regularly, as opposed to being from a more distant taxpayer. It is also possible that the gift being exchanged is not higher classroom effort for higher pay, but support from teacher unions to politicians in return for a pay increase. But our results do suggest that there was unlikely to have been a gift-exchange/reciprocity channel from higher pay to better job performance in this setting. They are also consistent with recent evidence suggesting that

perspective, we are more interested in whether such an expensive policy (which costs over 5 percent of the national budget) improved the effort of incumbent teachers and learning outcomes of their students through *any* combination of the posited mechanisms above. Our results suggest that even the composite effect of these mechanisms was negligible in this setting.

Our main contribution is to the literature on the personnel economics of the public sector. We are not aware of any experimental study to date on how a large unconditional salary increase affects the productivity of incumbent public employees. The most closely related paper is Mas (2006), which finds that police performance in New Jersey deteriorated when arbitrators awarded a lower pay increase than the one proposed by unions (relative to cases where the union proposal was accepted). One possible explanation for the seeming contrast with our results is gain-loss asymmetry around a reference wage point, with worker performance deteriorating in response to a pay cut relative to expectations but not improving in response to an unconditional increase in pay (see Koszegi and Rabin 2006 for theory, and Bewley 1999, and Kube, Maréchal, and Puppe 2013 for evidence). Since we study the effects of a large increase in salaries relative to the status quo, our results are not inconsistent with Mas's (2006) finding that employee performance does not improve with the gap between actual and expected pay when pay is above the reference point.

We also contribute to the literature on teacher pay and student performance. Our results are consistent with prior studies finding no correlation between increases in teacher pay and improved student performance in the US (Hanushek 1986; Betts 1995; Grogger 1996). However, these past results have been questioned for not having adequate exogenous variation in teacher pay, for failing to control for non-wage compensation and differences in local labor markets (Loeb and Page 2000), and for being based on changes in pay that may be too small to generate detectable impacts on outcomes (Dolton et al 2011). We are able to address all three of these limitations in our setting. In developing-country contexts, our results are consistent with other studies finding no correlation between teacher salaries in the public sector and their teaching effectiveness (Muralidharan and Sundararaman 2011, Bau and Das 2017), and with studies finding that contract teachers who are paid much lower salaries than civil-service teachers are no less effective (Muralidharan and Sundararaman 2013, Duflo, Dupas, and Kremer 2015, Bau and Das 2017).

any increases in worker productivity in response to an unconditional increase in pay may be short-lived (Gneezy and List 2006, Jayaraman, Ray, and de Véricourt 2016), or even non-existent (Kube, Maréchal, and Puppe 2013, Esteves-Sorenson 2017).

Our results *do not imply* that salary increases for public employees would have no positive impacts on service delivery in the long run through extensive-margin impacts. Dal Bo et al (2013) show that salary increases for public sector jobs in Mexico increased the quality of job applicants, and Ferraz and Finan (2011) find that higher wages for politicians in Brazil attracted more educated candidates and improved politician performance. Longer-term studies that include the extensive-margin effects of new teacher hiring have found a positive relationship between teacher salaries and student outcomes (Card and Krueger 1992a, 1992b, and Donohue, Heckman, and Todd 2002).

Rather, our results contribute to a more informed discussion on the cost-effectiveness of such a policy. Since the annual flow of new workers is low relative to the stock of existing workers, most extensive-margin benefits would accrue far in the future. In contrast, the costs of unconditional salary increases are incurred immediately (and are mostly driven by increased pay for incumbent workers). We show that at reasonable discount rates, the intensive-margin effects have a considerably greater weight than the extensive-margin effects in determining the present value of a policy of across-the-board pay increases. Thus, if there are no intensive-margin effects on productivity, implying that the extensive margin is the only channel for improved productivity, then our results suggest that across-the-board salary increases are a very inefficient way of improving education quality relative to alternate uses for public education funds (see calculations in section V).

Several global education policy reports recommend increasing teacher pay in low-income countries as a way to improve the motivation and performance of incumbent teachers (UNICEF 2011, UNESCO 2014). Following a similar set of arguments, the Government of Indonesia's publicly-stated rationale for the large salary increase included the hope that it would improve teacher morale, motivation, and job satisfaction, and thereby lead to increased teacher effort and student learning (see section II).⁵ Our results suggest that while the policy improved the welfare of incumbent teachers, it yielded no corresponding improvement in the learning of students taught by these teachers. Such evidence is especially relevant for improving policymaking in a public-sector

⁵ Note that politicians do not have to genuinely believe that higher salaries will raise teacher effort and effectiveness. They could also strategically claim to believe this because they need to present a plausible public-interest reason for raising teacher salaries in return for political support from teacher unions. In practice, both the "ideas" and the "interests" are likely to matter, but the "ideas" provide the *stated* rationale in both cases (see discussion in section II).

setting, where there is no market test of whether increasing employee salaries also increases productivity, and where unconditional pay raises are difficult to reverse.⁶

The rest of this paper is structured as follows: Section II describes the Indonesian education context, the teacher certification policy, and the mechanisms by which the policy could have improved teacher effort; section III describes our experiment (design, validity, and data collection); section IV presents our main results on the impacts on teacher welfare and student learning outcomes; section V interprets our results and discusses policy implications; section VI concludes. Tables A.1–A.8 and other appendices are available in an Online Appendix.

II. CONTEXT, POLICY REFORM, AND RATIONALE

Indonesia has one of the largest school education systems in the world, catering to a school-age population of over 50 million across 34 provinces and over 500 districts. The country consists of thousands of islands spanning over 3,000 miles from east to west (Figure I), making service delivery challenging. Promoting school education was historically a higher priority for Indonesia than for many other developing countries in South Asia and Africa, and primary school enrollment rates in Indonesia exceeded 90% by the early 1980s (World Bank EdStats Database). This priority on education was further formalized in 2000-02, when the new Indonesian constitution committed the government to spend at least 20% of its budget on education (a considerable increase from before).

Policy deliberations on the best way to spend these extra resources identified poor teacher quality and motivation as key limitations in the performance of the Indonesian education system. The ambitious education reforms of 2005 aimed to address this issue. The highlight of these reforms was the Teacher Law of 2005, which stipulated that teachers who met certain eligibility criteria (being a civil-service teacher, and holding either a four-year university degree, or a high rank in the civil service – typically obtained through a long tenure) and who successfully completed a

⁶ In contrast, Henry Ford's famous "five-dollar workday" led to a similar doubling in wages, but also led to sharp increases in worker productivity (Raff and Summers 1987). Indeed, it is unlikely that Ford would have continued paying high wages if productivity did not go up, whereas the Indonesian government spent billions of dollars on teacher salary increases and has continued doing so each year despite our results showing no impact on student learning outcomes. But, the government had no good way of knowing this *ex ante*, in the absence of evidence on the question.

certification process would receive a "professional allowance" (or "certification allowance") equal to 100% of their base pay (World Bank 2010, Chang et al. 2014).⁷

The certification process was initially meant to include a high-standards external assessment of teacher subject knowledge and pedagogical practice, with an extensive skill-upgrading component for teachers who did not meet these standards (featuring up to a year of additional training and tests). However, teachers' associations opposed the high-standards certification exams. Thus, by the time the final law and regulations were negotiated through the political and policymaking process, the quality-improvement stipulations had been highly diluted. They were replaced with a much weaker certification requirement that simply required teachers to submit a portfolio of their teaching materials and achievements. Even for those who did not pass the portfolio evaluation, just two weeks of additional training were required to attain certification. Thus, in practice, the certification process yielded a doubling of base pay with only a modest hurdle to be surmounted.⁸

Pre-reform teacher salaries in Indonesia were lower than teacher salary benchmarks in other Southeast Asian countries (which was part of the justification for the policy), but teachers were reasonably well paid even before the reform. Using representative household survey data from the 2012 Indonesian labor force survey (Sakernas, August 2012), we estimate that pre-reform teacher pay was at the 50th percentile of the college-graduate salary distribution. Civil-service teachers also enjoyed more generous benefits than equivalent workers in the private sector, and had high job security. Overall, teacher jobs were attractive even before the reforms, and quit rates were very low.

The reform led to a substantial increase in teacher salaries, moving teacher compensation from around the 50th percentile of the college-graduate salary distribution to the 90th percentile. This large salary increase was not conditional on teachers' subsequent effort or effectiveness, but instead depended only on a one-time determination that the teacher met some certification criteria. Hence, for all practical purposes, the policy can be considered as having resulted in an unconditional salary increase for eligible incumbent teachers. To the extent that undergoing the certification process

⁷ Note that the professional allowance was 100% of *base* pay, rather than of total pre-certification pay. Teachers often receive other allowances based on the location of their job posting and taking on additional tasks, and so the allowance increased total pay by 75% on average and by 65% for teachers who were eligible for certification (see Table IV).

⁸ Very few teachers entering the certification process failed it. Further, even those who failed the first attempt were all certified after a two-week training program, which mainly focused on helping teachers prepare the portfolios that would need to be submitted with the certification process (World Bank 2010).

increased teacher human capital, our estimates of the impact of certification will be an upper bound on the intensive-margin impacts of an unconditional increase in pay.

The decision to implement a large teacher pay increase was justified at least in part by the belief that higher pay would increase teacher motivation and effort. Indeed, the pay increase was widely referred to in policy documents as an "incentive", suggesting an implicit assumption by policy makers that there would be positive effects on teacher motivation and effort (see Chang et al. 2014). While standard economic models do not predict that workers will increase effort in response to an unconditional increase in pay, this belief is common in the global policy literature on teacher quality and was also reflected in the Indonesian policy discourse. For instance, UNESCO's flagship *Education for All Global Monitoring Report* claims that "[l]ow salaries reduce teacher morale and effort" and "teachers often need to take on additional work – sometimes including private tuition – which can reduce their commitment to their regular teaching jobs and lead to absenteeism" (UNESCO 2014). In Indonesia, one report claimed that "[l]ow pay is likely to be one of the main reasons why teachers perform poorly, and have low morale" (World Bank 2008), and another that "teachers often have a high rate of absenteeism because they take second jobs to make ends meet. This reality reduces their motivation and effectiveness in the classroom" (World Bank 2010).⁹

Appendix A illustrates how widespread this view is in policy circles by presenting a fuller list of quotes and extracts from prominent education policy documents in Indonesia and several other countries, which claim that increasing teachers' pay will increase their motivation and effort. In Appendix B, we formalize the economic arguments implied in these quotes from practitioners and present simple theoretical sketches of mechanisms by which teacher effort may increase in response to an unconditional pay increase and derive comparative statics. These include: (1) reciprocity and gift exchange in employment contracts; (2) a model in which effort on pro-social tasks like teaching is a normal good with a positive income elasticity; and (3) a model where the expected performance of teachers depends on their salary and where sanctions or rewards are provided through community and administrative monitoring based on performance relative to these expectations.¹⁰

⁹ The argument that higher teacher salaries can improve motivation and performance appears in the US literature as well. For instance, Hanushek, Kain, and Rivkin (1999) note that in addition to the attraction and retention channel, "Many influential reports and proposals advocate substantial salary increases as a means of attracting and retaining more talented teachers in the public schools *and encouraging harder work by current teachers*" (emphasis added).

¹⁰ We do not derive comparative statics for the "reduced shirking" channel of efficiency wages (Shapiro and Stiglitz 1984), because this is unlikely to apply in a public-sector setting where civil-service teachers are rarely fired.

Of course, the belief that unconditional pay increases would increase teacher morale, motivation, effort, and effectiveness is unlikely to have been the only reason for the policy change. As with any large policy change, the final Indonesian Teacher Law reflected a combination of "ideas" (people genuinely thinking that the salary increase would improve education outcomes), "interests" (teacher unions effectively advocating for their interests), and "institutions" (the spending floor on education in the new Indonesian Constitution allowed for a large increase in education spending). However, while "ideas" are only one part of this causal chain of action, they are especially important because they often provide the stated rationale for "interests". For instance, even if policy makers did not truly believe that the reform would improve education, and only wanted to reward teachers in return for political support, it may have made strategic sense for them to posit such effects as a plausible public-interest justification for the pay increase.

Thus, from a policy perspective, the private beliefs and publicly-stated rationales for the pay increase are less important than the fact that it was implemented, and was very expensive. Since the pay increase could have improved the effectiveness of incumbent teachers through several channels, the goal of our study is not to test any one channel of impact (which is not feasible); instead, we test whether the large pay increase helped to improve effort and productivity of incumbent teachers through *any* mechanism. Evidence on this question would inform future policy discussions on the cost effectiveness of large unconditional salary increases for incumbent civil-service employees.

III. EXPERIMENT DESIGN

III.A. Design, Sampling, and Implementation

Because of the large number of teachers covered, teacher access to the certification process was phased in for budgetary reasons. The budgetary restrictions meant that only around 10% of teachers were allowed to go through the certification process each year once the implementation of the certification process began in 2006. Each year, each district was allocated a quota that indicated how many of its teachers could start the certification process. The quota was typically allocated to teachers based on seniority, though districts had some discretion in this process. Once a teacher was in the process, he or she was practically guaranteed certification, as described above. Other eligible teachers had to wait in a certification queue, often for several years.

Our experimental design takes advantage of the phase-in procedure for teacher access to the certification process, and the existence of a certification queue. Rather than having teachers wait in this queue, the intervention aimed to allow all eligible but not yet certified teachers (whom we define as "target" teachers) in treatment schools to immediately access the certification process at the start of the experiment (in 2009). The experiment did not change any of the requirements of certification specified in the law and regulations, but simply allowed otherwise eligible teachers in treatment schools to enter the certification process early, rather than having to wait for a few more years. In other words, the experiment accelerated access to the certification and pay increase for teachers in treatment schools, but it did not change the underlying program in any way. The experimental protocol was implemented in close collaboration with the Ministry of National Education of the Government of Indonesia, where senior officials were committed to conducting a high-quality impact evaluation, and provided exemplary support in implementation.

We first identified a near-representative sample of 360 schools across 20 districts of Indonesia to comprise the universe of the study. We started with the 2006 national teacher census, which covered roughly 1,600,000 public primary and junior-secondary teachers across 454 districts. Districts that were too small, were too dangerous to visit, or were included in a parallel randomized evaluation were excluded,¹¹ leaving us with 383 districts in the sampling frame. These represented nearly 85% of the districts and over 90% of the population of Indonesia. From these, we randomly sampled 20 districts, stratified across the five major regions of the country, with more districts assigned to regions with a larger population. The list of districts sampled and the strata they represent are presented in Table A.1. A map of the sampled districts is presented in Figure I.¹²

¹¹ The district sampling for the two parallel sets of randomized evaluations was conducted using the same procedures, and so the 20 districts dropped on account of not wanting spillovers between the studies were also a representative sample. However, the second study (of a parallel initiative to set up teacher working groups) ended up not being implemented. Districts dropped for access and safety reasons had a much lower population on average.

¹² As the scale in Figure I indicates, the east-to-west distance spanned by Indonesia is greater than that of the continental United States, and our design imposed considerable logistical complexity. However, the resulting random assignment in a near-representative sample of schools provides greater external validity to our results. See Heckman and Smith (1995) for a discussion of the threats to external validity of experiments resulting from site-selection bias in experimental studies. Allcott (2015) provides evidence of such bias. The five major regions of Indonesia and the number of districts sampled in each of them (roughly proportional to population) were Java (10), Sumatra (5), Sulawesi (2), Eastern Indonesia (2), and Kalimantan (1).

Within each district, we stratified schools by the number of teachers, and sampled 12 primary and 6 junior secondary schools.¹³ Thus, the study universe consisted of a near-representative sample of 240 primary (grades 1-6) and 120 junior-secondary (grades 7-9) schools across 20 districts of Indonesia. From this sample, 80 primary and 40 junior-secondary schools were randomly assigned to "treatment" status, while the other 160 primary and 80 junior-secondary schools were assigned to a "business as usual" control group. Just like the sampling of schools, the randomization was also stratified by district, school type, and school size, and thus the design was identical across districts, with each district being a microcosm of the overall study.¹⁴

To implement the experiment, the Ministry of National Education sent letters to the District Education offices with a copy to the head teachers of treated schools informing them that all eligible teachers in the selected schools had been granted immediate access to the certification process, and informing them about the administrative steps they needed to take to enter the certification process (a translated copy of this letter is attached in Appendix C). To ensure that other teachers would have no incentive to transfer to treatment schools, only teachers who worked in the treatment schools at the start of the experiment were eligible for this immediate access.¹⁵ The budget for the extra certification "slots" needed for the experiment was provided through supplementary funds from the national government, and these slots were provided to districts over and above their regular quota. Thus, the experiment did not displace any other education spending in the districts from control to treatment schools; nor did it displace any otherwise eligible teacher from certification.

The research design did not create any change in the schools other than the additional quota allocation to treatment schools and the communication letter to head teachers of treatment schools. The teachers in control schools continued business as usual, with those who were eligible but not

¹³ We dropped the strata comprising schools with very large and very small number of teachers. If schools were too large, it would not have been feasible to test all the students in the school during the time that the enumerators would have in the school. If they were too small, they would not provide adequate power. Given that we find no evidence of heterogeneous effects as a function of the number of teachers in the school, our results are likely to be representative of all schools, even though the smallest and largest ones were not in the study universe.

¹⁴ Specifically, each of the 20 districts had 6 treatment schools (2 junior secondary; 4 primary) and 12 control schools (4 junior secondary; 8 primary). Schools were stratified into "triplets" based on size, and one school in each triplet was assigned to treatment status. Note that because the intervention was expensive, optimal sample allocation to maximize power yielded a larger control group than treatment group. All our estimating equations will include "district-triplet" fixed effects (since these are the strata within which we randomized treatment assignment).

¹⁵ The letter also promised accelerated access to non-eligible teachers in treatment schools once they met the eligibility criteria (point 2 in the letter). But as we show later, there was only limited impact on the certification rate of those who were not initially eligible, relative to the large impact on those who were eligible at the start of the experiment.

certified at the start of the study progressing through the certification process at the same rate as the rest of the country. Thus, our identifying variation comes from the sharp increase in the fraction of certified teachers in the treatment schools induced by the experiment, contrasted with the gradual, business-as-usual increase in the control schools.

The possibilities of spillovers to other schools were minimized by making sure that there was no public announcement of the additional quota: the eligibility for certification was communicated only to the head teacher and teachers in treatment schools through the letter that they received from the government. Further, within the treatment schools, the teachers who did not receive access to the certification process were those who were ineligible for certification in any case (by virtue of not being a college graduate or a civil-service teacher, for example). As a result, the experiment is less likely to have engendered resentment among non-target teachers in the school than in settings where the pay increases might have been seen as arbitrary. Thus, by conducting our study in a setting where the pay increases were in line with pre-announced policy criteria, we minimize the extent to which the intervention could be considered ad hoc or unsustainable.

III.B. Project Timeline and Data

The school year in Indonesia runs from July to May, and the study was carried out over three school years from 2009-10 to 2011-12. We refer to these three school years as Y1, Y2, and Y3 in the paper. The sampling and randomization of schools were conducted during the school holidays before Y1, and the government sent letters to treated schools informing them that all eligible teachers in these schools would be able to access the certification process at the start of Y1. The certification process (including preparing and submitting the application and teaching portfolio, having them evaluated, and receiving the certification) typically took one full school year, and teachers typically got "certified" by the end of Y1, and started receiving their certification allowance (equal to 100% of base pay) at the start of Y2 (the 2010-11 school year).

We carried out three waves of data collection, during which we interviewed head teachers, teachers, and students; and conducted independent tests of both teacher knowledge and student learning outcomes. The first wave was a baseline collected in October 2009. The baseline was deliberately conducted a few months into the school year (after the certification eligibility letters were sent to treatment schools) so that we could interview teachers to verify whether they had entered the certification process. The second wave of data was collected in April-May 2011, at the

end of 2 years of the project (Y2), and the third wave was collected in April-May 2012, at the end of 3 years (Y3).¹⁶ Figure II shows the project timeline for the intervention and data collection.

We collected data on school facilities, finances, and other school-level data from head-teacher interviews. Teacher interviews included questions on demographics, experience, pay, outside jobs, income (from teaching and other sources), and job satisfaction. We used a combination of school and teacher interviews to map teachers to specific classrooms and subjects (which will not be needed for the school-level ITT estimates, but will be needed for the IV estimates of the impact of being taught by a certified teacher). Students in all schools were tested using multiple-choice tests of math, science, and Indonesian, and students in junior secondary schools were also tested in English. The tests also included a short demographic survey to collect basic information on household assets from students.

III.C. Validity of the Experimental Design

The randomization was successful in ensuring that treatment and control schools were similar prior to the experiment. There was no significant difference between treatment and control schools on school-level variables such as the number of students, teachers, or class size (Table I - Panel A). There were also no significant differences in student test scores across treatment and control schools on test scores in any subject (math, science, Indonesian, or English) or on an index of household assets (Table I - Panel B).¹⁷ The differences in means in column (3) include "district-triplet" fixed effects, since these triplets are the strata within which we randomized treatment assignment, and the stratum fixed effects will be included in our estimating equation for treatment effects.

Teacher characteristics were also similar across treatment and control schools. There were no significant differences on most teacher-level variables, including teachers' own test scores, their certification status, their base pay, and the incidence of holding an outside job (Table II: Columns 1-3). The only major difference is that, as expected, teachers in treatment schools were 32 percentage points more likely to have entered the certification quota. This difference confirms that the

¹⁶ Since the certification process took one year, the first year in which target teachers in treatment schools would have received the additional allowance was the second year of the project. We felt it was highly unlikely that there would be any impact at the end of Y1 (since teachers in treatment schools would not have received any additional payments at this point). Thus, given the high costs of surveys across the Indonesian islands, we did not collect data at the end of Y1.

¹⁷ Note that the randomization (and communication to "target" teachers) was carried out before the baseline survey and hence the randomization could not be balanced ex ante on these variables. Thus, it is reassuring to see that treatment and control schools were balanced on observables.

experiment successfully led to many more teachers in treatment schools getting access to the certification process.

We see the impact of the treatment even more clearly in Table II: Columns 4-6, which are restricted to the target teachers who were "eligible but not certified" in either the treatment or control schools at the start of the study. In this group, 73% of teachers in treatment schools were in the certification quota; whereas in the control schools, the rate was only 18% (indicating the rate at which "target" teachers would have gotten certified in the absence of the experiment). We do observe small differences in a few other teacher characteristics that are attributable to random sampling variation. However, the magnitudes of these differences are small, especially when compared to the differences in the fraction admitted to the certification quota. To control for these differences, we also report results from a differences-in-differences specification when we look at impacts at the teacher level.¹⁸

We also test for differential attrition and entry of students over the period of the study. Table A.2 shows the different cohorts in our study, the years in which they were tested, and which cohorts are in our estimation sample at different points of the study. We find that there is no differential attrition among students who were in our baseline test and who continue to be in our estimation sample over time (Table A.4 – Panel A), and also that there is no difference in attrition rates across treatment and control groups as a function of baseline test scores (Panels B and C). Finally, we find that the treatment did not induce any compositional changes in incoming student cohorts over time as measured by a household asset index (Table A.5).

IV. RESULTS

IV.A. First-Stage

The time path of the fraction of teachers in treatment and control schools who had entered the certification process over the three years of the study is shown in Figure III. Three points are

¹⁸ Teachers in treatment schools were slightly more likely to have a bachelor's degree, but slightly less likely to have a senior civil-service rank. These factors offset each other in determining certification eligibility, and we see no difference in the fraction of certification-eligible teachers across treatment and control schools (56% vs. 57% - columns 1 and 2). There are small differences in pre-certification pay, but these are less than 5% of the value of the certification pay. The significance of these small differences is attributable to the very small standard errors obtained from including the stratum fixed effects (the differences are mostly not significant without the stratum fixed effects).

noteworthy. First, there was no difference between treatment and control schools in the rate of teacher certification before the start of the experiment in 2009. Second, the intervention introduced a sharp increase in the fraction of teachers admitted to the certification process in treatment schools in 2009, even as the trend in control schools remained constant. Third, the gap in fraction of admitted teachers narrowed over time, as the eligible teachers in the control schools gained access to the certification process at a "business as usual" rate. Thus, the difference in the fraction of teachers admitted to the certification process across treatment and control schools is higher at the time of the baseline survey (Y0) than at the end of Y2 and Y3.

As described earlier, teachers entered the certification process at the start of each school year, completed the process over the course of the year, got certified by the end of the year, and started receiving their payments at the start of the next year. Thus, at the time of the baseline there was no difference between treatment and control schools in the fraction of teachers who were certified or who had received the extra certification allowance. However, both indicators had increased sharply by the end of Y2 and Y3 (Figure IV).

Table III - Panel A shows the differences in Figures 3 and 4, along with tests of equality. In the first year, the share of teachers in treatment schools who had entered the certification process was 33 percentage points higher than (or more than double) that in the control group, while no difference had yet appeared in the fraction certified or paid the certification allowance. At the end of Y2 and Y3, the difference in the fraction of teachers who had entered the certification process falls to 17 and 7 percentage points respectively (since the control schools catch up over time). At the end of Y2 (Y3), the fraction of teachers in treatment schools who report being certified is 23 (14) percentage points higher, and the fraction who report being paid the certification allowance is 28 (23) percentage points higher.

Note that the difference in the fraction of teachers who are paid their certification allowance is higher than the difference in the fraction who are certified (at the end of both Y2 and Y3). This is as expected; many eligible teachers in the control schools would have entered the certification process at the start of Y2 and Y3 and then been certified at the end of Y2 and Y3 respectively, but would have started getting paid their allowances only at the start of the next school year. These teachers will therefore report being certified but will not yet have started getting paid their allowance at the time of the Y2 and Y3 surveys, respectively. On the other hand, teachers in treatment schools who gained access to the certification process at the start of Y1 will have completed certification by the

end of Y1, and started getting paid their allowances in Y2.¹⁹ Since most of the posited mechanisms by which the pay increase would be expected to improve teacher effort and student outcomes are based on teachers actually receiving the extra pay, the most relevant metric of the "effective difference" between treatment and control schools for our study is the difference in the fraction of teachers who have been paid their certification allowance.

We present the corresponding figures for the "target" teachers—those who were eligible but not certified at the start of the study—in Table III – Panel B. As expected, the differences are more pronounced for this group. The target teachers in treatment schools are 54 percentage points more likely to have entered the certification process at the time of the baseline survey. At the end of Y2 (Y3), they are 43 (24) percentage points more likely to be certified, and 54 (45) percentage points more likely to have been paid their certification allowance (Table III - Panel B).

Finally, we present the corresponding figures for the "non-target" teachers, who were not eligible for certification at the start of the experiment in Table A.3. The experiment also aimed to provide accelerated access to certification to teachers in treatment schools who became eligible for certification in later years (as seen in point 2 in the letter in Appendix C). However, as Table A.3 shows, only very few of the teachers who were not eligible at the start of the experiment get "certified and paid" during the study (2% after Y2 and 3% after Y3). Our estimates of intent-to-treat (ITT) effects at the school-level will include these teachers, and our instrumental variable (IV) estimates will focus on teachers who were eligible at the start of the study (where the first-stage is the highest).

IV.B. Teacher-Level Outcomes

Table IV reports the impact of the experiment on teachers in treated schools after two and three years. Columns 1-6 report impacts for all teachers (which documents the first stage for the school-level ITT effects), and columns 7-12 report impacts for target teachers (which corresponds to the first stage for the IV estimates of the impact of being taught by a teacher who received the pay

¹⁹ Thus, the variation in the difference between treatment and control groups across measures reported in Table III reflects variation in the year of entry into the certification process and the time lag in the process. Once we control for year of entry into certification, the difference between treatment and control schools in the fraction of teachers who are certified and the fraction who are "certified and paid" is the same.

increase). We report both simple differences (with stratum fixed effects) and difference-in-difference estimates that adjust for differences in baseline value (whenever these are available).

We find that the accelerated access to the certification process and the additional allowance had several positive impacts on teachers that persisted both two years and three years into the study. At the end of Y2 (Y3), teachers in treatment schools received 112% (72%) more certification pay and 19% (15%) more total pay compared to those in control schools.²⁰ They were also 15% (12%) more likely to report being satisfied with their total income, 18% (16%) less likely to report facing financial problems and stress, 18% (18%) less likely to be holding a second job, and spent 19% (16% - not significant) less time working on second jobs (Table IV – columns 1-6).

As we would expect, the impacts are stronger within the universe of target teachers. At the end of Y2 (and Y3), target teachers in treatment schools received 274% (103%) more certification pay and 31% (23%) more total pay than those in control schools. Note that the certification allowance was 100% of base pay for teachers, but that in practice, the increase over their total pre-certification pay was around 65-75% because the total pay (prior to certification) included allowances in addition to their base pay.²¹ Compared to their peers in control schools, target teachers in treatment schools were also 28% (20%) more likely to report being satisfied with their total income, 27% (31%) less likely to report facing financial problems and stress, 19% (12%) less likely to be holding a second job, and spent 22% (10%) less time working on second jobs at the end of Y2 (Y3) (Table IV – Columns 7-12).²²

²⁰ These figures are presented in percentage changes relative to the mean in the control group. Table IV presents the changes in percentage points. Calculations in the text use the difference-in-difference estimates when available, and the simple difference estimates otherwise. As an illustration, columns 1 and 3 of Table IV show that the mean certification pay in the control group at the end of Y2 was 0.57M IDR (Indonesian Rupiah) and that the treatment raised this by 0.64M IDR, yielding an increase of 0.64/0.57; this is the 112% figure reported in the text.

²¹ It is easy to back this out from the numbers in Tables III and IV. In the sample with all teachers, we see in Table III that 27% of teachers in the control group had been paid the certification allowance in Y2, and see in Table IV that the mean certification pay in the control group was 0.57M IDR. Thus, the average certification pay among the teachers who were receiving it was 0.57M/0.27, which is 2.11M IDR. This is, as it should be, a 100% increase over the mean base pay of 2.08M IDR in the control group (Table IV – column 1). Base pay plus allowances equals 2.85M IDR, so certification pay was 74% of pre-certification pay (2.11M/2.85M). The calculation can also be done with the target teachers, where we see that the average certification pay conditional on receiving it in Y2 was 0.38M/0.18 in the control group, which is also 2.11M IDR. But since other allowances for senior civil-service teachers were higher, the total pre-certification pay for the "target" teachers was 3.25M IDR. Thus, target teachers received a 65% increase (2.11M/3.25M) in their total pay upon certification.

²² Results on incidence of second jobs and time spent on second jobs are often not significant in Y3 (likely reflecting the weaker first stage of the treatment in Y3 as certification rates in the control schools catch up over time).

Since eligible teachers in control schools would also become eligible for certification over time, our experiment did not induce a doubling in permanent income. Rather, it *accelerated* a permanent doubling of base pay, and increased lifetime income for target teachers by 2 to 3 years of base pay. Further, while eligible teachers in control schools may have been able to anticipate their future increase in income, credit constraints may have limited the extent to which they could borrow against future income. Thus, the effects we report above on increased job satisfaction, reduced financial stress, and reduced outside jobs should be interpreted as the result of an increase in 2 to 3 years of permanent income, as well as the liquidity effects of receiving the extra income on hand.²³

Overall, the teacher pay increase induced by our experiment was successful in achieving the stated objectives of the certification policy regarding teachers' financial situation, job satisfaction, and ability to better focus on teaching by reducing the need to hold outside jobs. However, we find little evidence to suggest that teachers in treatment schools put in greater effort in response to this pay increase. We find no difference between treatment and control schools on teacher test scores or the likelihood of pursuing further education, suggesting that teachers did not use the extra time available for their primary teaching job to upgrade their skills. We also find no difference in self-reported absence rates in three out of four comparisons in Table IV (last row, columns 3, 6, 9, and 12), suggesting that teacher effort may not have changed much in treated schools.²⁴

Nevertheless, as per the mechanisms described in section II (and Appendices A and B), the reduced financial stress, reduced incidence of second jobs, and increased job satisfaction and motivation could have led to an improvement in teacher effort in the classroom, and effectiveness as measured by student learning outcomes. We test for this possibility in the next section.

IV.C. Student Outcomes

Intent-to-Treat (ITT) Estimates. Since the randomization was conducted at the school level, we first present school-level intent-to-treat estimates. These estimates quantify how student learning in

²³ Note also that there is no reason to expect the experiment to affect the teachers in the control schools. They already knew about the policy, and had access to the certification process in exactly the same way as they would have had without the experiment. The experiment only accelerated the pay increase for teachers in treated schools, but did not change any way in which control schools experienced the larger certification reform (the certification process was unchanged during the period of the experiment).

²⁴ The results in the last row of Table IV – Column 9 suggest that teacher absence was lower among target teachers in treated schools (who are the group we would most likely see an effort response for). However, these results are based on self-reports of absence, which limits our confidence in inferring impacts on teacher effort. As a result, our primary outcome of interest is student learning, which we measure through independently administered tests.

a school responds to a sharp increase in the fraction of the school's teachers who have received a large unconditional increase in pay. Our main estimating equation takes the form:

$$(1) \quad T_{ijks}(Y_n) = \beta_0 + \beta_1 \cdot \overline{T_{ijks}}(Y_0) + \beta_j \cdot T_{ijks}(Y_0) + \beta_2 \cdot Treatment_k + \beta_{Z_{ST}} \cdot Z_{ST} + \varepsilon_{ijks}$$

The dependent variable of interest is T_{ijks} , which is the normalized test score of student i on subject s , where j, k , denote the grade and school respectively. $T(Y_0)$ indicates the baseline tests, while $T(Y_n)$ indicates a test at period Y2 or Y3. Including the normalized baseline test score improves efficiency, due to the autocorrelation between test scores across multiple periods.²⁵ We also include a set of stratum fixed effects (Z_{ST}), to account for the stratification of the randomization. Finally, we include the mean normalized baseline test scores across all students in the school for the corresponding grade and subject ($\overline{T_{ijks}}$), which further increases efficiency (Altonji and Mansfield 2014). The main estimate of interest is β_2 , which provides an unbiased estimate of the impact of being in a "Treatment" school (the intent-to-treat or ITT estimate), since schools were assigned to treatment status by lottery.

Table V presents these ITT estimates pooled across schools and subjects; we see that there was no impact on test scores of being in a treated school, even though teacher salaries and satisfaction had gone up substantially. The pooled effects across subjects and school types have a point estimate of -0.01σ at the end of Y2 and 0.01σ at the end of Y3. These zero effects are precisely estimated; the small standard errors of 0.025σ provide us adequate power to detect effects as low as 0.05σ at the 5% level. Thus, not only are the point estimates close to zero, but we can also reject effect sizes greater than 0.04σ at the end of Y2 and greater than 0.06σ at the end of Y3. Table A.6 presents results individually for each subject, by school type (primary and junior secondary) and at the end of Y2 and Y3 (Panel A and B); the results show that there is no effect on test scores in any subject at either of the two time periods (columns 1-4).

Figure V presents quantile treatment effects of being in a treatment school, by plotting student test scores at each percentile of the control and treatment school test score distributions after Y2 and

²⁵ As we show in Table A.2, some of the cohorts included in our analysis did not have a baseline test. We set the normalized baseline score to zero for these students (similarly for students who were absent for the baseline test but are present in the Y2 and/or Y3 tests) and include a dummy variable in equation (1) that takes the value 1 when the lagged test score is missing and 0 when it is present. We also allow the coefficient on the lagged test score to vary by grade.

Y3 (left hand side plots). We see that the treatment effects are not only zero on average, but also cannot be statistically distinguished from zero at any part of the test score distribution. On the right-hand side, we present the corresponding first-stage quantile plots, which show the number of years that a student at each quantile of the test-score distribution spent with a certified teacher in a treatment and control school. The figure makes clear that students at every percentile of the test-score distribution after Y2 and Y3 experienced a significant increase in their exposure to a certified teacher, but that nevertheless there was no impact on learning outcomes.

One possible concern in interpreting our school-level ITT estimates is that the estimated zero effects could reflect a combination of positive effects on students of target teachers (who may be motivated to increase effort by the pay raise) and negative effects on students taught by "non-target" teachers (especially those who were not eligible for certification), who may have reduced effort in response to the perceived "unfairness" of not receiving the certification allowance.²⁶ We test for this possibility by decomposing the impact on mean test-scores shown in Table V into test-score impacts on students taught by target teachers and those taught by non-target teachers (across treatment and control schools). We present the results in Table VI, for both Y2 and Y3.

For the Y2 data, we consider whether a student was taught by a target teacher in Y2 (since none of the target teachers would have been paid the certification allowance in Y1), and test separately for treatment effects on students taught by target and non-target teachers (Table VI – column 1). We see that there is no effect on test scores of students in treatment schools taught by either type of teacher relative to the control schools (point estimates are zero), and cannot reject equality of test scores of students taught by target and non-target teachers in treatment schools.²⁷

For the Y3 data, we consider the four possible combinations of teacher type that a student could have had in Y2 and Y3 (target – target; target – non-target; non-target – target; and non-target – non-target) and again find no significant different in test-score outcomes across these categories between treatment and control schools. When we focus on the most extreme comparison – students taught by a target teacher in both Y2 and Y3, compared with those taught by a non-target teacher in

²⁶ As described earlier, this was unlikely because the experiment did not change any of the certification norms in the law, and thus there is no reason for non-eligible teachers to feel such resentment. But we still test for this possibility.

²⁷ The table separately reports outcomes for the small fraction of students (around 5% of observations) for whom we are not able to verify the target status of their teacher (reported as “no match between student and teacher”).

both Y2 and Y3 – we still find no evidence that the former did better in treated schools (Table VI – column 2).

Instrumental Variable (IV) Estimates. The ITT estimates presented above are at the school level, and are based on a 29- (24-) percentage-point increase in the fraction of "certified and paid" teachers in the treatment schools at the end of Y2 (Y3) (Table III – Panel A). To estimate the direct impact of being taught by a "certified and paid" teacher, we instrument for being taught by a certified teacher using the random assignment of treatment across schools. Specifically, we aim to estimate:

$$(2a) \quad T_{ijks}(Y_2) = \beta_0 + \beta_1 \cdot \overline{T_{ijks}}(Y_0) + \beta_j \cdot T_{ijks}(Y_0) + \beta_2 \cdot \text{Certified}_{ijks}(Y_2) + \beta_{Z_{ST}} \cdot Z_{ST} + \varepsilon_{ijks}$$

$$(2b) \quad T_{ijks}(Y_3) = \beta_0 + \beta_1 \cdot \overline{T_{ijks}}(Y_0) + \beta_j \cdot T_{ijks}(Y_0) + \beta_2 [\text{Certified}_{ijks}(Y_3) + Y \cdot \text{Certified}_{ijks}(Y_2)] + \beta_{Z_{ST}} \cdot Z_{ST} + \varepsilon_{ijks}$$

where the coefficient of interest is β_2 , which estimates the impact on student test scores for each year of being taught by a *Certified* teacher (with the additional pay), and the rest of the variables are defined as in Eq. (1).

One technical consideration in estimating Eq. (2b) is the issue of test-score decay (or incomplete persistence) over time. Estimates from several settings suggest that there is considerable annual decay in test scores, with the persistence parameter Y (estimated as the coefficient on the lagged test score in a standard value-added model) typically being around 0.5 (Andrabi et al. 2011). It is not possible to consistently estimate the persistence parameter and a treatment effect for later years of the treatment at the same time (see the discussion in Andrabi et al. 2011 and Muralidharan 2012). We therefore estimate Eq. (2b) for a range of values of Y and present the resulting estimates of β_2 , along with standard errors, in Table VII. The estimates with $Y = 0$ correspond to complete decay of any test score gains in a year by the end of the next year, while those with $Y = 1$ correspond to complete persistence. Based on several prior studies, our preferred estimates assume $Y = 0.5$.

The main threat to interpreting these estimates as the annual impact of being taught by a certified teacher is the possibility of endogenous re-assignment of certified teachers within treatment schools to potentially weaker students. We test for this in Table A.7 and find that there is no significant difference across treatment and control schools in the probability of a student being assigned to target teachers as a function of assets or test scores during either Y2 or Y3 (Table A.7 –

Panel A). We also find no difference in the probability of students being assigned to a target teacher as a function of their incoming test scores (based on comparing Y0 scores in Y2 and Y2 scores in Y3), and whether they are above or below the median asset ownership (Table A.7 – Panels B-E).²⁸

Table VII presents IV estimates of the impact of being taught by a certified teacher for the full sample of students, as well as for the sample of students taught by target teachers (which will give us more precise IV estimates, since the first stage is more powerful in this case). Focusing on students who were taught by target teachers, we can reject a positive effect greater than 0.07σ at the 95% level in the Y2 data. In the Y3 data, our preferred estimate is the one where the sample includes students who were taught by a target teacher in either Y2 or Y3, and we find that we can reject a positive effect greater than 0.1σ at the 95% level.²⁹

Finally, we examine heterogeneity of treatment effects as a function of several school-level characteristics, including the fraction of target teachers, the number of target teachers, mean student affluence, measures of school size, mean baseline test scores, and find no evidence of any heterogeneous effects (Table A.8). Thus, the increase in teacher pay in treated schools had no impact on student test scores, either in aggregate or in any subset of the data.

V. COST EFFECTIVENESS AND POLICY IMPLICATIONS

Before discussing cost effectiveness, we note that teacher salary increases do not represent a social cost, because they are a transfer from taxpayers to teachers. The social cost of the program is the deadweight loss of raising tax revenue for the increased salaries, combined with the cost of implementing the certification program. However, developing countries typically face hard budget constraints because of a limited ability to run deficits, and so the cost of the policy may best be thought of as the opportunity cost of potentially higher-return public spending that was crowded out.³⁰ To simplify our analysis, we limit the use of this "opportunity cost" framework to other

²⁸ Note that we test for differential assignment of students to target teachers as a function of the household asset index because we do not have baseline test scores for many of the cohorts in our final estimation sample.

²⁹ We also show the ITT effects for each estimation sample in Table VII to enable a comparison between ITT and IV estimates. These are almost identical since outcomes are similar across students taught by target and non-target teachers (as seen in Table VI).

³⁰ In principle, governments should be able to borrow to finance any project that has a higher rate of return than the cost of borrowing. In practice, financial markets find it difficult to evaluate the quality of public spending and impose a sovereign-risk interest-rate penalty when fiscal deficits exceed a threshold. Thus, in practice, choosing one form of public spending will reduce the fiscal space for other policies, which motivates our "opportunity cost" approach.

education expenditure. We assume that there is a fixed education budget, and compare this program to other education interventions that could have been implemented with the same resources.

Since the salary doubling had no impact on test scores of students taught by incumbent teachers, it is clear that the policy was not cost-effective as a way of improving the quality of education for current students.³¹ Thus, the case for across-the-board teacher salary increases as a policy option for improving student learning would have to rely exclusively on longer-run impacts—the possibility that, over time, education quality could improve as higher-quality candidates enter the teaching profession. We provide suggestive estimates on the potential magnitude of this effect below.

Using data on teacher subject-knowledge test scores matched to student value-added from the data set used in this study, de Ree (2016) estimates that a 1σ increase in teacher test scores predicts a 0.175σ /year increase in their effectiveness as measured by student value-added (the estimates are from page 28 of de Ree 2016). So, if we assume that the doubling of pay attracted and led to the selection of teachers who have 1σ better subject test scores than the current stock of teachers, the extensive-margin effect would be to improve student test scores by around 0.175σ /year in steady state after all current teachers have been replaced.³²

Thus, in the long-run steady state, the policy may yield an increase in student test scores of 0.175σ /year through extensive-margin effects at a cost of USD 138 per student per year.³³ However, other salary-related interventions in developing countries have led to comparable increases in learning at much lower cost. For instance, a program that provided individual performance-based bonus pay to teachers in India achieved student test-score gains of 0.15σ /year (averaged across math and language) at an annual cost of only about USD 4 per student, including implementation costs

³¹ In contrast, several other interventions have been able to achieve substantial test-score gains for existing students in developing countries (see Glewwe and Muralidharan 2016 for a review). Thus, if the policy goal of the government was to improve learning outcomes of current students, then it is likely that one or more of these other programs could have been implemented in Indonesia with the resources spent on the salary increases and delivered greater test-score gains.

³² The assumption is not unrealistic in theory since the pay increase moved teacher salaries from the 50th to the 90th percentile of the distribution of college-graduate salaries (a pay increase of over 1σ if salaries are normally distributed). However, in practice, it is very optimistic since it assumes that the teacher selection process would also be modified to *select* the higher-ability candidates who may be attracted to teaching by the higher pay, which was not the case in the status quo. For instance, as of 2012 (6 years after the reform), nearly 50% of recently recruited teachers (between 24-30 years of age) did not have a bachelor's degree, despite there being no shortage of college graduates with a teaching degree, suggesting that status quo teacher hiring did not select the most qualified candidates (World Bank 2015).

³³ Costs were calculated by taking the monthly certification allowance (2.11M IDR – from section IV.2), multiplying this by 12 and the average number of teachers (9.3, from Table I), and dividing by the average number of children in a school (190, from Table I), using a 9000 IDR/US dollar exchange rate from the period of the experiment (2009-2012). Because it assumes no growth in real teacher salaries over time, this is a conservative estimate of costs.

(Muralidharan and Sundararaman 2011).³⁴ Expressed as a fraction of teacher base pay (since India and Indonesia have different levels of GDP per capita), the performance-pay program in India cost 6% of base pay (3% each for bonus and implementation costs), while the across-the-board salary increase in Indonesia cost 100% of base pay. Thus, even when considering the potential long-term steady state benefits of the pay increase on learning outcomes, it is likely that an alternative policy of performance-linked pay increases would be much more cost-effective.

Three further considerations suggest that across-the-board salary increases are even less cost-effective from a social welfare perspective. First, such increases result in large and *immediate* fiscal costs by increasing pay levels of incumbent workers. Thus, the short- and medium-term benefits (net of costs) depend largely on the magnitude of the intensive-margin effects (which we show to be zero), while most of the extensive-margin effects accrue only far in the future, as older cohorts of teachers retire and newer cohorts join the teacher work force. In Appendix D, we show that at a discount rate of 7% (which is the interest rate on 10-year Indonesian government bonds), the intensive-margin effects of a policy of raising salaries across the board have a weight three times greater than that of the extensive-margin effects in calculating the present value of the policy.

Specifically, if E_i and E_e are the steady-state annual intensive- and extensive-margin effects on student learning, respectively, we show that the present value of the discounted stream of benefits from the policy is equal to $(E_i \times 15) + (E_e \times 5)$. We also show that if the annual steady-state cost of the salary increase is C , then the present value of the discounted stream of costs will be $(C \times 15)$.³⁵ In other words, if E_i is zero (as we find), and the discount rate is 7%, then the present discounted value of the stream of costs is over 15 times higher than the annual figure (since the costs start

³⁴ Incentive treatments cost up to rupees 10,000 per school. Per-student costs are obtained by dividing by average student in school (113), and then using an exchange rate of 44 rupees to the dollar (in the years of the experiment 2005-2007), yielding a cost of 2 US dollars per student. The authors conservatively estimate the cost of implementing the program as equal to the costs of the bonuses; including the implementation cost would double the per-child cost to 4 USD per student, which is the figure we use.

³⁵ The presented discounted value of a continuous stream of annual costs C , is equal to $\{C/(1-\delta)\}$, where δ is the discount factor, which is equal to $1/(1+r)$, where r is the discount rate. Thus, if r is 7%, then $\{1/(1-\delta)\}$ is 15.28, yielding the estimate in the text. The multiple for the intensive-margin effect (E_i) is analogous since this effect also starts immediately. However, the multiple for the extensive-margin effect (E_e) is lower because these benefits only phase in over time (see calculations in Appendix D). Note also that from a public budgeting perspective, both C and E should be expressed in dollars to determine whether an investment has a positive rate of return. In practice, the mapping from test-score gains to wage gains (and hence economic return) is not well documented in most countries. We therefore follow the spirit of the discussion in the opening paragraph of this section: we think about E in terms of standard deviations of test scores, and we focus on the relative cost-effectiveness of different policies aimed at improving test scores.

immediately), while the present discounted value of the stream of benefits is only 5 times higher (since the gains from E_e appear only in the longer run). The calculation also highlights the importance of the intensive-margin effects for the present value calculation, and shows how our results inform cost-effectiveness calculations. If E_i were positive instead of zero, the present value of the salary increase could be much higher.

Second, even if such an increase raises the quality of new entrants into the teaching profession, it is not obvious that this will improve social welfare, because that talent would be *displaced* from other sectors in the economy (unlike policies that improve the effectiveness of *existing* teachers). While it is possible that the social returns of attracting more talented individuals to teaching may be higher than the costs to the sector they are displaced from, there is no evidence of this. Further, since public-sector management quality and productivity is typically lower than that of the private sector (Bloom and Van Reenen 2010), it is possible that higher-quality human capital may be less productive in the public sector, and that the displacement reduces aggregate output.³⁶

Third, an alternative policy that links at least some of the pay increases to performance is likely to not only yield positive intensive-margin effects, but also be more effective on the extensive margin. This is because increasing the spread of worker pay to more closely reflect their productivity is also likely to attract higher-ability candidates, compared with an across-the-board increase in salaries on a compressed schedule with no links to performance (Lazear 2000). In the context of education, Muralidharan and Sundararaman (2011b) find that teachers in India who are *ex ante* more willing to accept a mean-preserving spread in pay linked to their performance are the ones who are more effective *ex post*. Thus, while increasing teacher compensation across the board may have some positive long-term effects on education outcomes through its effects on teacher quality, our results and the discussion above suggest that there may be much more cost-effective ways of improving education outcomes.

³⁶ For instance, Schuendeln and Playforth (2014) present evidence from India suggesting that educated workers prefer to join the government sector (which has high wages and high private returns) even though the social returns of the government sector are low. More recently, Bau and Das (2017) show that there is no correlation between teacher value-added and teacher pay in the public sector in Pakistan, while there is a positive correlation between the two in private schools, suggesting that the private sector is able to manage employees better (by rewarding performance). Finally, another under-appreciated cost of salaries in the public-sector being high relative to market norms is that it could induce corruption in recruitment into government jobs and induce negative selection of candidates who are willing to pay bribes to obtain well-paid lifetime employment (see Muralidharan 2016 for evidence and discussion).

VI. DISCUSSION AND CONCLUSION

This paper has offered new evidence on a key question in public-sector personnel economics: How does a large, unconditional increase in salary affect the job performance of incumbent employees? This is an important policy question because most of the cost of unconditional salary increases is devoted to paying higher salaries for these incumbents. The value of evidence on this question is especially important in public-sector contexts, where there is no market test of whether such an increase is a cost-effective way of improving the effort and effectiveness of employees.

We answer this question with a large-scale randomized experiment in the context of a policy change in Indonesia that led to a permanent doubling of base teacher salaries. The experiment was implemented successfully, leading to a large increase in teacher incomes in treated schools. It also substantially improved the intermediate variables through which policymakers hoped that the increase in salary would lead to better education quality: teachers in treated schools were significantly more likely to be satisfied with their income, significantly less likely to report financial stress, and less likely to hold a second job than teachers in control schools.

Yet despite this improvement in teachers' pay and satisfaction, we find no effect on teacher effort towards upgrading their own skills, no consistent evidence of changes in self-reported teacher attendance, and no effect on the ultimate outcome of student learning. The test-score impact of being in a treated school is close to zero, and we can rule out effects as small as 0.05σ at the 95% level in treated schools. Similarly, the test-score impact of being taught by a certified teacher who had received the pay increase was also close to zero, and we can rule out positive test score effects larger than 0.1σ at the 95% level. Thus, it appears that the large increase in teacher salaries was mostly a transfer to teachers without any corresponding improvement in productivity.

Advocates of higher pay for teachers frequently assert that it would improve the motivation, effort, and effectiveness of existing teachers (as discussed in Appendices A and B). These ideas influence the broader public discourse on education, contributing to expensive policy changes of the sort implemented in Indonesia. Our results suggest that this hypothesis is not supported by the evidence. Further, while our study was not designed to test specific mechanisms (such as gift-exchange and reciprocity, or more effective supervision) by which unconditional salary increases may improve the effort and effectiveness of incumbent employees, our results suggest that *none* of these posited channels applied in our setting of civil-service workers with high job security.

These results are directly relevant to policy debates—around the world, and especially in developing countries— regarding whether across-the-board salary increases for teachers (and other public-sector employees) are a cost-effective strategy for improving their productivity and the quality of service delivery more broadly. While such pay increases could improve the quality of entrants into teaching and improve student learning in the longer run, these extensive-margin effects will appear only after many years, while the costs are borne immediately (mainly for spending on incumbent workers). Our calculations show that if the intensive-margin effects are zero, leaving the extensive margin as the only channel of impact, then unconditional salary increases are unlikely to be a cost-effective policy option for improving the quality of service delivery.³⁷

More broadly, our results are consistent with a growing body of evidence showing that wages of public-sector workers in developing countries are typically not correlated with productivity (see Das et al. 2016 for evidence from public-sector healthcare workers, and Muralidharan and Sundararaman 2011 and Bau and Das 2017 for evidence from education). Whereas much of the existing evidence is correlational, we provide experimental evidence that unconditional pay increases do not increase public-sector worker productivity. Conversely, the fact that the policy was implemented is consistent with the hypothesis that public-sector compensation policy does not reward productivity; this may help explain why management quality is lower in public organizations than in private firms, which are significantly more likely to compensate service providers for greater productivity (Bloom and Van Reenen 2010, Bau and Das 2017).

Compared to these changes in level of compensation, reforms to the structure of public-sector worker compensation (especially using performance-linked bonuses) appear more promising as a strategy for improving service delivery (Muralidharan and Sundararaman 2011). However, implementing such reforms at scale in public-sector settings is much more challenging. Given the centrality of front-line worker effort and productivity to service delivery in developing countries, there are likely to be large returns to future research on the personnel economics of the public sector, and specifically on the effectiveness (or lack thereof) of policies to improve public-sector worker productivity.

³⁷ One policy option that mitigates this problem is to only have the higher salaries apply to new recruits (thereby obtaining extensive margin benefits without the intensive margin costs on incumbent workers that may not raise effort and productivity), but this is likely to be considered unfair and be politically difficult to implement.

WORLD BANK
UC SAN DIEGO, NBER, BREAD, AND J-PAL
UNIVERSITY OF AMSTERDAM, VU AMSTERDAM, TINBERGEN INSTITUTE, AND AIGHD
WORLD BANK

REFERENCES

- Akerlof, George A, "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97 (1982), 543-569.
- Allcott, Hunt, "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130 (2015), 1117-1165.
- Altonji, Joseph G., and Richard K. Mansfield, "Group-Average Observables as Controls for Sorting on Unobservables When Estimating Group Treatment Effects: The Case of School and Neighborhood Effects," (National Bureau of Economic Research, Inc, NBER Working Papers: 20781, 2014).
- Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc, "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics," *American Economic Journal: Applied Economics*, 3 3 (2011), 29-54.
- Bau, Natalie, and Jishnu Das, "The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers," (2017).
- Betts, Julian R, "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *The Review of Economics and Statistics*, (1995), 231-250.
- Bewley, Truman F, *Why Wages Don't Fall During a Recession* (Cambridge, MA: Harvard University Press, 1999).
- Bloom, Nicholas, and John Van Reenen, "Why Do Management Practices Differ across Firms and Countries?," *Journal of Economic Perspectives*, 24 (2010), 203-224.
- Card, David, and Alan B Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *The Journal of Political Economy*, 100 (1992), 1-40.
- , "School Quality and Black-White Relative Earnings: A Direct Assessment," *The Quarterly Journal of Economics*, 107 (1992), 151-200.
- Chang, Mae Chu, Samer Al-Samarrai, Andrew B. Ragatz, Joppe de Ree, Sheldon Shaeffer, and Ritchie Stevenson, *Teacher Reform in Indonesia: The Role of Politics and Evidence in Policy Making* (Washington, DC: World Bank, 2014).
- Clements, Benedict, Sanjeev Gupta, and Masahiro Nozaki, "What Happens to Social Spending in IMF-Supported Programmes?," *Applied Economics*, 45 (2013), 4022-4033.
- Dal Bó, Ernesto, Frederico Finan, and Martín A Rossi, "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, 128 (2013), 1169-1218.
- Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan, "Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India," *American Economic Review*, 106 (2016), 3765-3799.
- de Ree, Joppe, "How Much Teachers Know and How Much It Matters in Class," (World Bank, 2016).
- Dolton, P., O. D. Marcenaro-Gutierrez, L. Pistaferri, and Y. Algan, "If You Pay Peanuts Do You Get Monkeys? A Cross-Country Analysis of Teacher Pay and Pupil Performance," *Economic Policy*, (2011), 5-55.
- Donohue III, John J, James J Heckman, and Petra E Todd, "The Schooling of Southern Blacks: The Roles of Legal Activism and Private Philanthropy, 1910–1960," *The Quarterly Journal of Economics*, 117 (2002), 225-268.

- Duflo, Esther, Pascaline Dupas, and Michael Kremer, "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools," *Journal of Public Economics*, 123 (2015), 92-110.
- Esteves-Sorenson, Constanca, "Gift Exchange in the Workplace: Addressing the Conflicting Evidence with a Careful Test," *Management Science*, (Forthcoming).
- Falk, A., "Gift Exchange in the Field," *Econometrica*, 75 (2007), 1501-1511.
- Ferraz, Claudio, and Frederico Finan, "Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance," UC Berkeley, 2011.
- Finan, Frederico, Benjamin A Olken, and Rohini Pande, "The Personnel Economics of the State," in *Handbook of Field Experiments*, Esther Duflo, and Abhijit Banerjee, eds. (North Holland, 2017).
- Glewwe, P, and K Muralidharan, "Improving Education Outcomes in Developing Countries," *Handbook of the Economics of Education*, 5 (2016), 653-743.
- Gneezy, U., and J. A. List, "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74 (2006), 1365-1384.
- Government of India, "Report of the Sixth Central Pay Commission," (2008).
- , "Report of the Seventh Central Pay Commission," (2015).
- Grogger, Jeff, "School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond," *The Review of Economics and Statistics*, (1996), 628-637.
- Hanushek, Eric A, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of economic literature*, 24 (1986), 1141-1177.
- Hanushek, Eric A., John F. Kain, and Steven G. Rivkin, "Do Higher Salaries Buy Better Teachers?," (National Bureau of Economic Research, Inc, NBER Working Papers: 7082, 1999).
- Heckman, J. J., and J. A. Smith, "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9 (1995), 85-110.
- International Monetary Fund, "Managing Government Compensation and Employment--Institutions, Policies, and Reform Challenges," (Washington, DC: International Monetary Fund, 2016).
- Jalal, Fasli, Muchlas Samani, Mae Chu Chang, Ritchie Stevenson, Andrew B Ragatz, and Siwage D Negara, "Teacher Certification in Indonesia: A Strategy for Teacher Quality Improvement," (Jakarta: World Bank, 2009).
- Jayaraman, Rajshri, Debraj Ray, and Francis de Véricourt, "Anatomy of a Contract Change," *The American Economic Review*, 106 (2016), 316-358.
- Kőszegi, Botond, and Matthew Rabin, "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121 (2006), 1133-1165.
- Kube, Sebastian, Michel Andre Marechal, and Clemens Puppe, "Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment," *Journal of the European Economic Association*, 11 4 (2013), 853-870.
- Lazear, Edward, "Performance Pay and Productivity," *American Economic Review*, 90 (2000), 1346-1361.
- Loeb, Susanna, and Marianne E. Page, "Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-Pecuniary Variation," *Review of Economics and Statistics*, 82 3 (2000), 393-408.
- Mas, Alexandre, "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 121 (2006), 783-821.
- Mullis, Ina VS, Michael O Martin, Pierre Foy, and Alka Arora, *Timss 2011 International Results in Mathematics* (2012).
- Muralidharan, Karthik, "Long-Term Effects of Teacher Performance Pay," (UC San Diego, 2012).
- , "A New Approach to Public Sector Hiring in India for Improved Service Delivery," *India Policy Forum 2015-16*, 12 (2016), 187-225.
- Muralidharan, Karthik, and Venkatesh Sundararaman, "Teacher Opinions on Performance Pay: Evidence from India," *Economics of Education Review*, 30 (2011), 394-403.

- , "Teacher Performance Pay: Experimental Evidence from India," *Journal of political Economy*, 119 (2011), 39-77.
- OECD, "Pisa 2012 Results in Focus: What 15-Year-Olds Know and What They Can Do with What They Know," (2013).
- Raff, Daniel M. G., and Lawrence H. Summers, "Did Henry Ford Pay Efficiency Wages?," *Journal of Labor Economics*, 5 4 (1987), S57-86.
- Sabnavis, Madan, and Anuja Shah, "Impact of 7th Pay Commission," (CARE Ratings, 2015).
- Schündeln, Matthias, and John Playforth, "Private Versus Social Returns to Human Capital: Education and Economic Growth in India," *European Economic Review*, 66 (2014), 266-283.
- Shapiro, C., and J. E. Stiglitz, "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74 (1984), 433-444.
- UNESCO, *Teaching and Learning: Achieving Quality for All. Efa Global Monitoring Report 2013/14*. (Paris: UNESCO, 2014).
- UNICEF, *Teachers: A Regional Study on Recruitment, Development and Salaries of Teachers in the Cecis Region* (Geneva: UNICEF, 2011).
- VSO, "Teaching Matters: A Policy Report on the Motivation and Morale of Teachers in Cambodia," (London: VSO International, 2008).
- Webb, Richard, and Sofia Valencia, "Human Resources in Public Health and Education in Peru," in *A New Social Contract for Peru: An Agenda for Improving Education, Health Care, and the Social Safety Net*, Daniel Cotlear, ed. (2006).
- World Bank, *Spending for Development: Making the Most of Indonesia's New Opportunities. Indonesia Public Expenditure Review* (Washington, DC: World Bank, 2008).
- , *Transforming Indonesia's Teaching Force* (Jakarta: World Bank, 2010).

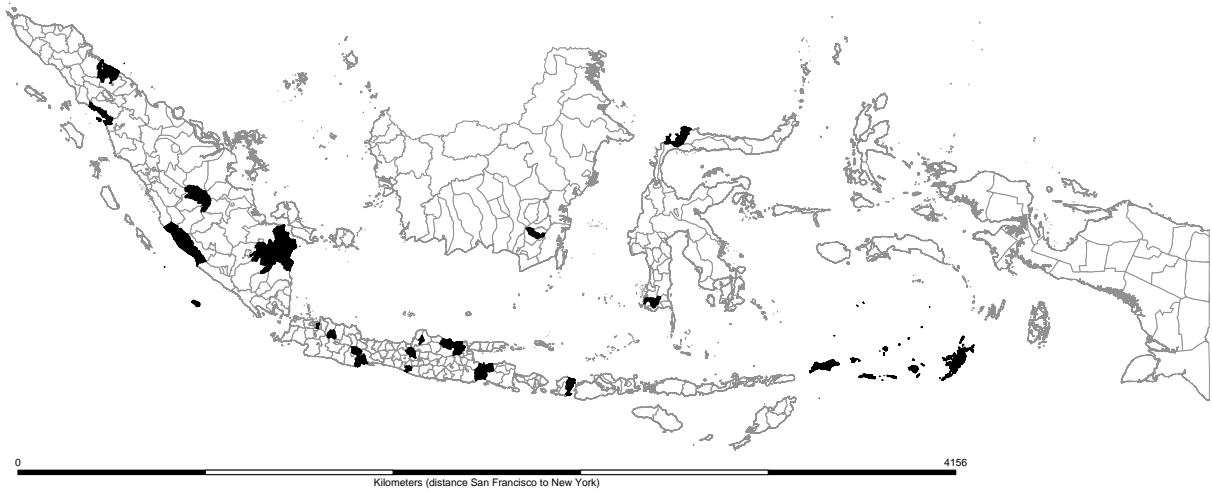


FIGURE I
Map of Indonesia with 20 Selected Districts

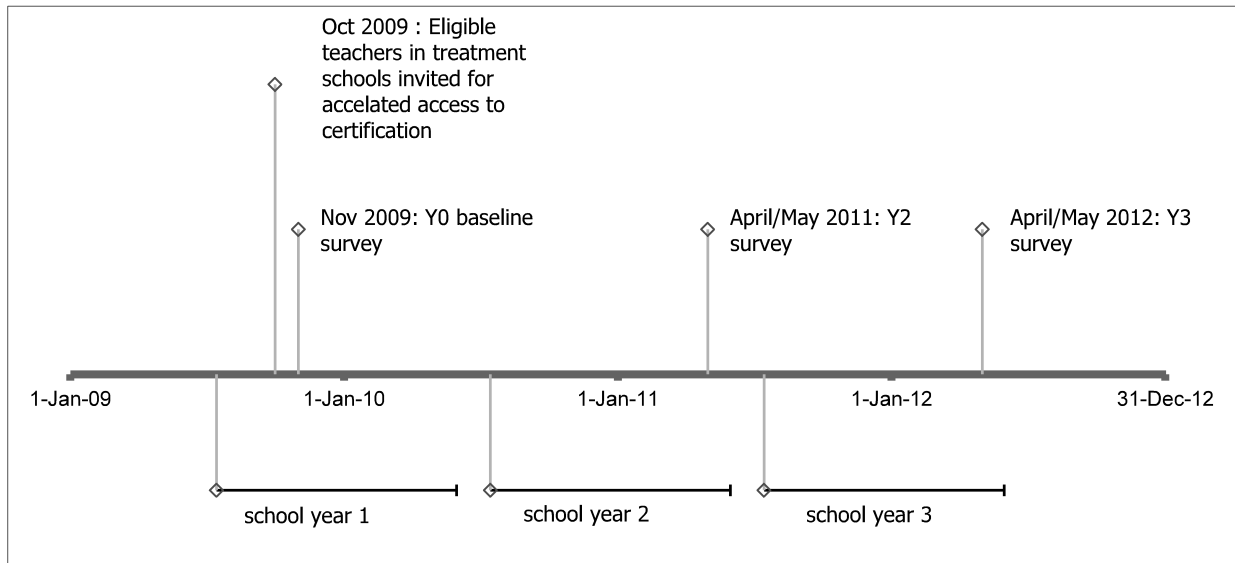


FIGURE II
Project Timeline

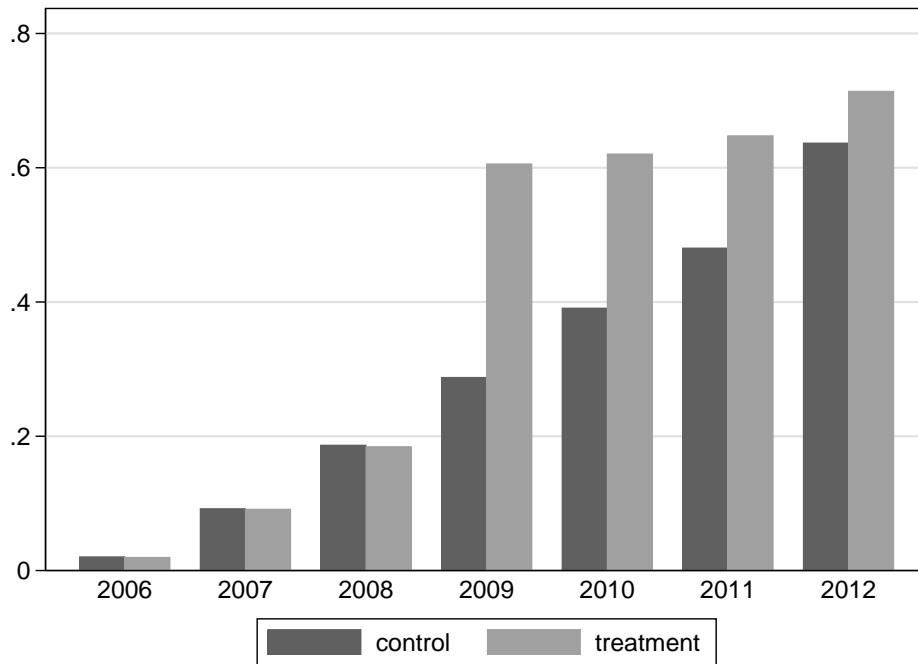


FIGURE III

Fraction of Teachers Admitted to the Certification Process, At, or Before the Indicated Year

Notes: Teachers were admitted to the certification process at different points in time. The first batch of teachers was admitted in 2006. The intervention took place in 2009, which created a difference between treatment and control schools in terms of the fraction of teachers admitted to the certification program. The bars represent fractions of teachers who were admitted to the certification program at, or before the indicated year. For example, around 60% of teachers in treatment schools were admitted to the certification program in the year 2009 or before, against roughly 30% in control. We use baseline data to construct the 2006, 2007, 2008, 2009 bars, Y2 data to construct the 2010 and 2011 bars, and Y3 data to construct the 2012 bar.

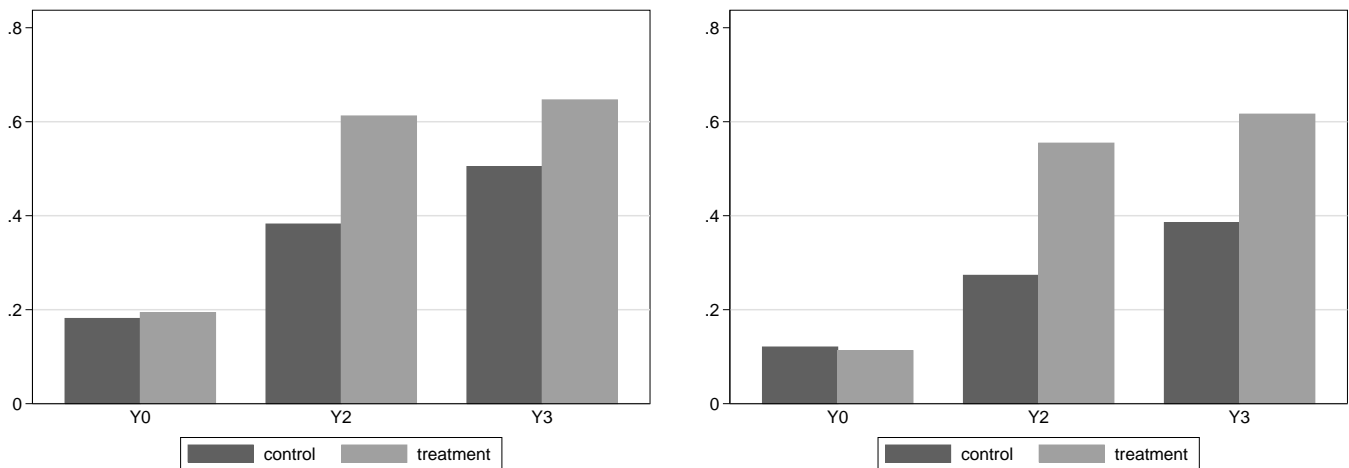


FIGURE IV

Completing the Certification Process (Left) and Being Paid the Certification Allowance (Right)

Notes: The left panel presents the fraction of teachers who completed the certification process. The right panel presents the fraction of teachers who completed the certification process and were paid the certification allowance.

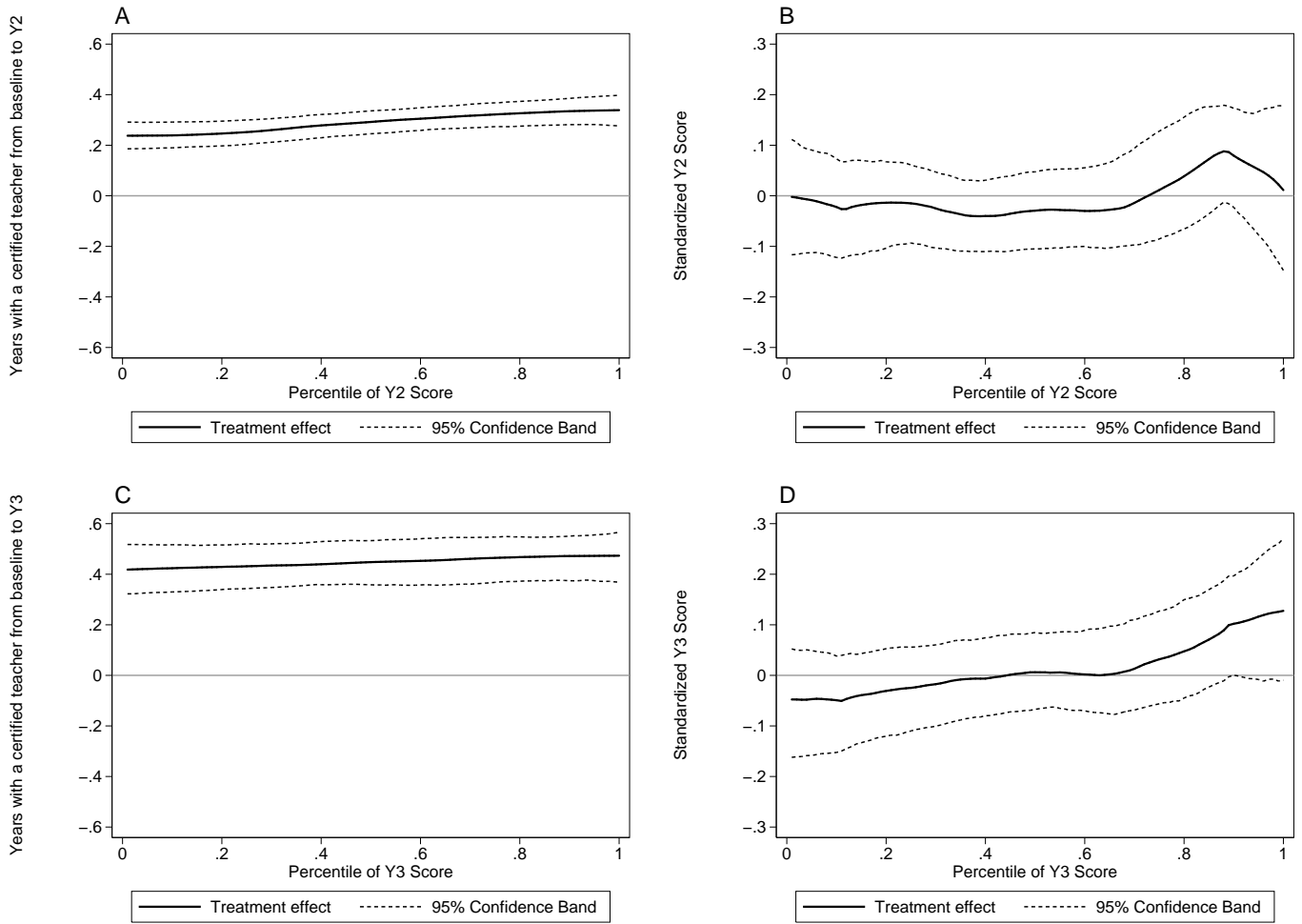


FIGURE V

Quantile Treatment Effects [Panel B (Y2), Panel D (Y3)] and Quantile First Stage [Panel A (Y2), Panel C (Y3)]

Notes: The nonparametric plots are constructed as follows: First, the outcome variable is regressed on a full set of district-stratum dummy variables, the school averaged baseline score [which is set to zero when it is not observed], and a dummy variable indicating observations for which the school-averaged baseline test scores are not observed. The residuals of this regression are linked to the percentiles using a local polynomial smoother. Percentiles on the horizontal axis are constructed separately for the treatment and control group. The confidence bands are estimated using a bootstrap method, and allow for residual dependence within schools.

TABLE I
BALANCE TESTS ON SCHOOL- AND STUDENT-LEVEL VARIABLES AT BASELINE

	(1)	(2)	(3)
Panel A: Balance test on school-level variables			
	Treatment	Control	Difference (F.E)
Number of classes per school	8.89 [4.88]	8.32 [4.49]	0.57 (0.35)
Number of students per school	190.85 [133.80]	184.49 [135.32]	6.36 (10.41)
Class size	20.60 [6.76]	20.99 [7.16]	-0.39 (0.64)
Number of teachers per school	9.35 [5.20]	9.07 [4.59]	0.27 (0.36)
Observations	120	240	
Panel B: Balance test on student-level variables			
	Treatment	Control	Difference (F.E)
Raw math score	0.41 [0.23]	0.40 [0.23]	-0.00 (0.01)
Raw science score	0.51 [0.21]	0.52 [0.21]	-0.00 (0.01)
Raw Indonesian score	0.58 [0.21]	0.59 [0.20]	-0.01 (0.01)
Raw English score	0.40 [0.18]	0.39 [0.17]	0.01 (0.01)
Student assets index	0.55 [0.24]	0.53 [0.24]	0.00 (0.01)
Observations	20,970	41,192	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table compares average baseline values between treatment and control groups based on a regression model that includes district-triplet fixed effects, which are the strata used for randomization. Within-group standard deviations are reported in brackets in columns (1) and (2). School-level clustered standard errors of the estimated difference between treatment and control is reported in parentheses in column (3). For the student asset index we calculate the fraction of the following seven items that are available in the household of the student: television, fridge, mobile phone, bicycle, motor bike, car, computer.

TABLE II
BALANCE TESTS ON TEACHER-LEVEL VARIABLES

	(1)	(2)	(3)	(4)	(5)	(6)
	All teachers			Target teachers only		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Raw (fraction correct) test score	0.56 [0.16]	0.56 [0.16]	0.00 (0.01)	0.55 [0.17]	0.56 [0.17]	-0.01 (0.01)
Eligible but not certified at baseline (i.e., Target)	0.56 [0.50]	0.57 [0.50]	-0.01 (0.02)	1.00 [0.00]	1.00 [0.00]	-0.00 <i>d.n.a.</i>
Already certified at baseline	0.19 [0.39]	0.18 [0.38]	0.02 (0.01)	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a.</i>
Not eligible for certification at baseline	0.25 [0.43]	0.25 [0.43]	-0.00 (0.01)	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a.</i>
Bachelor's degree	0.62 [0.49]	0.59 [0.49]	0.04*** (0.02)	0.69 [0.46]	0.65 [0.48]	0.06*** (0.02)
High rank (rank IV) in civil service	0.41 [0.49]	0.44 [0.50]	-0.03 (0.02)	0.48 [0.50]	0.51 [0.50]	-0.04** (0.02)
Certified and paid the certification allowance	0.11 [0.32]	0.12 [0.33]	-0.01 (0.01)	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a.</i>
Base pay (in mil. IDR)	1.87 [0.83]	1.92 [0.80]	-0.05 (0.03)	2.02 [0.73]	2.07 [0.69]	-0.07** (0.03)
Other allowances (in mil. IDR)	0.53 [0.34]	0.54 [0.33]	-0.02 (0.01)	0.55 [0.31]	0.59 [0.31]	-0.04*** (0.01)
Certification pay (in mil. IDR)	0.21 [0.59]	0.22 [0.60]	-0.01 (0.02)	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a.</i>
Second job	0.34 [0.47]	0.34 [0.47]	-0.00 (0.02)	0.33 [0.47]	0.35 [0.48]	-0.02 (0.02)
Hours worked on second job (last week)	3.50 [8.04]	3.40 [7.69]	0.12 (0.29)	3.18 [6.99]	3.40 [7.48]	-0.19 (0.34)
Started or completed the certification process	0.61 [0.49]	0.29 [0.45]	0.33*** (0.02)	0.73 [0.45]	0.18 [0.39]	0.54*** (0.03)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table compares average values between treatment and control groups for all teachers (left panel) and target teachers (right panel). Estimates are based on a regression model that includes district-triplet fixed effects, which are the strata used for randomization. Within-group standard deviations are reported in brackets in columns (1), (2), (4) and (5). School-level clustered standard errors of the estimated mean differences between treatment and control are reported in parentheses in columns (3) and (6).

TABLE III
FIRST STAGE PROCESS - TEACHER LEVEL

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Y0			Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Panel A: All teachers									
Started or completed the certification process	0.61 [0.49]	0.29 [0.45]	0.33*** (0.02)	0.65 [0.48]	0.48 [0.50]	0.17*** (0.02)	0.71 [0.45]	0.64 [0.48]	0.07*** (0.02)
Completed the certification process	0.19 [0.40]	0.18 [0.39]	0.01 (0.01)	0.61 [0.49]	0.38 [0.49]	0.23*** (0.02)	0.65 [0.48]	0.50 [0.50]	0.14*** (0.02)
Certified and paid the certification allowance	0.11 [0.32]	0.12 [0.33]	-0.01 (0.01)	0.55 [0.50]	0.27 [0.45]	0.29*** (0.02)	0.62 [0.49]	0.39 [0.49]	0.24*** (0.02)
Panel B: Target teachers only									
	Y0			Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Started or completed the certification process	0.73 [0.45]	0.18 [0.39]	0.54*** (0.03)	0.86 [0.35]	0.55 [0.50]	0.32*** (0.03)	0.90 [0.30]	0.82 [0.38]	0.08*** (0.02)
Completed the certification process	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a.</i>	0.81 [0.39]	0.39 [0.49]	0.43*** (0.03)	0.86 [0.35]	0.62 [0.48]	0.24*** (0.03)
Certified and paid the certification allowance	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a.</i>	0.72 [0.45]	0.18 [0.38]	0.54*** (0.03)	0.83 [0.38]	0.40 [0.49]	0.45*** (0.03)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table compares average values between treatment and control groups for all teachers (Panel A) and target teachers (Panel B). Estimates are based on a model that includes district-triplet fixed effects, which are the strata used for randomization. Within-group standard deviations are reported in brackets in columns (1), (2), (4), (5), (7) and (8). School-level clustered standard errors of the estimated mean differences between treatment and control are reported in parentheses in columns (3), (6) and (9).

TABLE IV
TEACHER-LEVEL IMPACT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All teachers						Target teachers only					
	Y2			Y3			Y2			Y3		
	Control mean	ITT (simple diff.)	ITT (diff. in diff.)	Control mean	ITT (simple diff.)	ITT (diff. in diff.)	Control mean	ITT (simple diff.)	ITT (diff. in diff.)	Control mean	ITT (simple diff.)	ITT (diff. in diff.)
Standardized test scores	0.01 [0.99]	0.00 (0.05)	0.04 (0.05)	0.01 [0.99]	-0.06 (0.05)	-0.04 (0.05)	0.01 [0.98]	0.03 (0.06)	0.09* (0.05)	0.05 [0.98]	-0.08 (0.06)	-0.05 (0.06)
Bachelor's degree	0.68 [0.47]	0.04*** (0.01)	0.01 (0.01)	0.73 [0.44]	0.05*** (0.02)	0.01 (0.01)	0.72 [0.45]	0.05** (0.02)	0.00 (0.01)	0.75 [0.43]	0.05** (0.02)	0.01 (0.02)
Pursuing further education	0.18 [0.39]	-0.01 (0.01)		0.16 [0.37]	-0.03** (0.01)		0.08 [0.28]	0.01 (0.01)		0.08 [0.26]	0.03* (0.02)	
Second job	0.32 [0.47]	-0.06*** (0.02)	-0.06*** (0.02)	0.27 [0.44]	-0.05** (0.02)	-0.04** (0.02)	0.31 [0.46]	-0.06** (0.02)	-0.06** (0.02)	0.25 [0.43]	-0.03 (0.03)	-0.03 (0.02)
Hours worked on second job last week	2.98 [7.41]	-0.56** (0.26)	-0.46* (0.24)	2.52 [6.15]	-0.40 (0.26)	-0.28 (0.25)	2.63 [6.27]	-0.69** (0.32)	-0.58* (0.30)	2.27 [5.76]	-0.23 (0.35)	-0.19 (0.33)
Base pay (in mil. IDR)	2.08 [0.94]	-0.09** (0.04)	-0.01 (0.01)	2.59 [0.74]	-0.04 (0.03)	0.00 (0.01)	2.35 [0.75]	-0.08** (0.04)	-0.01 (0.02)	2.77 [0.59]	-0.04 (0.03)	-0.00 (0.02)
Other allowances (in mil. IDR)	0.77 [0.75]	-0.01 (0.02)	0.02 (0.02)	0.62 [0.64]	-0.05* (0.03)	-0.04 (0.03)	0.90 [0.79]	-0.03 (0.03)	0.03 (0.03)	0.69 [0.59]	-0.08** (0.04)	-0.03 (0.04)
Certification allowance (in mil. IDR)	0.57 [0.97]	0.55*** (0.04)	0.64*** (0.04)	0.88 [1.23]	0.49*** (0.06)	0.63*** (0.06)	0.38 [0.83]	1.03*** (0.05)	1.04*** (0.05)	0.92 [1.25]	0.94*** (0.08)	0.95*** (0.08)
Total Pay (in mil. IDR)	3.41 [1.97]	0.44*** (0.08)	0.66*** (0.05)	4.29 [1.95]	0.43*** (0.09)	0.64*** (0.07)	3.62 [1.58]	0.92*** (0.08)	1.11*** (0.06)	4.48 [1.70]	0.89*** (0.10)	1.03*** (0.08)
<i>Baseline controls</i>		no	yes		no	yes		no	yes		no	yes

TABLE IV
TEACHER-LEVEL IMPACT
(CONTINUED)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All teachers						Target teachers only					
	Y2			Y3			Y2			Y3		
	Control mean	ITT (simple diff.)	ITT (diff. in diff.)	Control mean	ITT (simple diff.)	ITT (diff. in diff.)	Control mean	ITT (simple diff.)	ITT (diff. in diff.)	Control mean	ITT (simple diff.)	ITT (diff. in diff.)
Financial problems	0.50 [0.50]	-0.09*** (0.02)		0.56 [0.50]	-0.09*** (0.02)		0.48 [0.50]	-0.13*** (0.02)		0.51 [0.50]	-0.16*** (0.03)	
Satisfied with total income	0.60 [0.49]	0.09*** (0.02)		0.60 [0.49]	0.07*** (0.02)		0.60 [0.49]	0.17*** (0.02)		0.65 [0.48]	0.13*** (0.02)	
Absent from school at least once last week	0.14 [0.34]	-0.00 (0.01)	-0.02 (0.01)	0.13 [0.33]	0.01 (0.01)	-0.00 (0.01)	0.12 [0.32]	-0.03* (0.02)	-0.04** (0.02)	0.10 [0.31]	0.00 (0.02)	-0.01 (0.02)
<i>Baseline controls</i>		no	yes		no	yes		no	yes		no	yes

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports intent-to-treat effects at the teacher level. Column (1), (4), (7) and (10) report averages values in the control group for reference. Columns (2), (5), (8) and (11) reports treatment effects based on a model that includes district-triplet fixed effects, which are the strata used for randomization. Columns (3), (6), (9) and (12) report treatment effects based on a model that includes district-triplet fixed effects and baseline values as controls. Within-group standard deviations reported in brackets in columns (1), (4), (7) and (10). School-level clustered standard errors of the estimated treatment effects are reported in parentheses in columns (2), (3), (5), (6), (8), (9), (11) and (12). Empty cells in columns (3), (6), (9) and (12) correspond to variables for which we do not have baseline values.

TABLE V
INTENT-TO-TREAT EFFECTS ON STUDENT TEST SCORES

	(1) Y2	(2) Y3
Treatment effect	-0.005 (0.024)	0.010 (0.026)
Observations	279,066	274,993
R^2	0.28	0.24

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports intent-to-treat effects on student-level test scores. Estimates are reported separately for Y2 and Y3 data. Test score data are constructed by standardizing by subject-grade-year (so that mean and variance in the control group are 0 and 1 respectively), then stacked so that the unit of observation is student-subject-year. These test scores are then regressed on a dummy variable indicating a treatment school. The estimated parameter on the treatment indicator is reported in column (1) and (2). The regression model further includes district-triplet fixed effects (the strata used for randomization), baseline standardized student-level test scores, baseline standardized averaged school-level test scores. For observations for which baseline test scores are not observed, the baseline values are set to zero. Two dummy variables indicating observations for which individual baseline test scores or school averaged baseline scores are not observed, are also included in the regression model. Weights are applied to scale the student-subject level data back to the level of the student. School-level clustered standard errors reported in parentheses.

TABLE VI
INTENT-TO-TREAT EFFECTS ON STUDENT TEST SCORES, BY TARGET STATUS OF TEACHERS

	(1) Y2	(2) Y3
<i>Conditional treatment effect for students with:</i>		
Target teacher in Y_2	-0.004 (0.026)	
Non-target teacher in Y_2	-0.001 (0.030)	
Target in Y_2 and Y_3		-0.023 (0.033)
Target in Y_2 and non-target in Y_3		0.055 (0.041)
Non-target in Y_2 and target in Y_3		0.020 (0.039)
Non-target in Y_2 and Y_3		0.029 (0.036)
<i>no match between student and teacher</i>	-0.092 (0.056)	-0.045 (0.049)
H_0 : causal parameters are the same (p -value)	0.91	0.26
H_0 : causal parameters are the same and equal to zero (p -value)	0.99	0.37
Total student-subject observations	279,066	274,993

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports intent-to-treat effects on student-level test scores, conditional on the target-status of teachers. The regression model used is the same as the regression model used for the table (V) results, except that the dummy variable indicating a treatment school is interacted with a variable measuring the target status of a student's teacher. For Y_2 outcome data (column (1)) we present causal parameters for students with target teachers in Y_2 , and for students who do not have target teachers in Y_2 . For Y_3 outcome data (column (2)) there are four categories, as indicated. For a minority of students we could not match students to teachers (for example when teachers were absent during the field visit). For column (1) results, about 5.6% of the student-test-level observations were not matched to teachers in Y_2 . For column (2) results, about 12.2% of the student-level observations were not matched to teachers in both Y_2 and Y_3 . The treatment effects for the subgroup of students that are not matched to teachers are not statistically significant, suggesting that the inability to match students to teachers does not cause biases. School-level clustered standard errors reported in parentheses.

TABLE VII

IV RESULTS, MEASURING ANNUAL STUDENT TEST SCORE GAINS DUE TO CERTIFICATION AND PAYING CERTIFICATION ALLOWANCE TO TEACHERS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Y2				Y3					
	ITT	IV	Upperbound 95% conf. interval	<i>N</i>	ITT	IV ($\Upsilon = 0.0$)	IV ($\Upsilon = 0.5$)	IV ($\Upsilon = 1.0$)	Upperbound 95% conf. interval	<i>N</i>
Full Y_2 sample	-0.005 (0.024)	-0.016 (0.078)	0.137	267,792						
Target teacher in current year	-0.014 (0.025)	-0.025 (0.046)	0.065	138,749						
Full Y_3 sample					0.014 (0.026)	0.060 (0.105)	0.039 (0.070)	0.029 (0.052)	0.176	241,438
Target teacher in current year					-0.012 (0.029)	-0.028 (0.068)	-0.021 (0.050)	-0.016 (0.040)	0.077	116,490
Target teacher in current OR in previous year					0.005 (0.026)	0.013 (0.072)	0.009 (0.047)	0.006 (0.035)	0.101	151,788
Target teacher in current AND in previous year					-0.036 (0.032)	-0.084 (0.076)	-0.051 (0.046)	-0.037 (0.033)	0.040	54,463

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports IV estimates where *effective* number years with a certified teacher since baseline, is instrumented with the dummy variable indicating a treatment school. The variable *effective* number years with a certified teacher takes into account that there is imperfect persistence of effects across time, see Equation (2b) in the main text. Control variables used in the regression model are the same as those used for the Table V and VI results. The IV model is estimated on different subsets of the data (in rows). We consider the full Y_2 and Y_3 samples, as well as selected groups who were more strongly affected by the intervention. Columns (3) and (9) report upperbounds on the 95% confidence intervals. For the calculation of this value we use the estimates reported in column (2) and (7) respectively. School-level clustered standard errors reported in parentheses.

**DOUBLE FOR NOTHING? EXPERIMENTAL EVIDENCE ON AN UNCONDITIONAL
TEACHER SALARY INCREASE IN INDONESIA**

JOPPE DE REE
KARTHIK MURALIDHARAN
MENNO PRADHAN
HALSEY ROGERS

ONLINE APPENDICES

**Appendix Tables A.1 – A.8
Appendices A-D**

TABLE A.1
STRATA AND SAMPLED DISTRICTS

Strata	Sampled districts
Eastern Indonesia (Maluku and Papua)	Maluku Tenggara Barat
Nusa Tenggara	Lombok Timur
Western Java	Ciamis, Jakarta Timur, Purwakarta
Central Java	Bantul, Kudus, Semarang
Eastern Java + Bali	Lamongan, Lumajang, Probolinggo, Tuban
Kalimantan	Hulu Sungai Selatan
Sulawesi	Gowa, Toli Toli
Northern Sumatra	Deli Serdang, Tapanuli Tengah
Western Sumatra	Tebo
Southern Sumatra	Bengkulu Utara, Ogan Ilir

Notes: Regions (the strata) are approximate geographic descriptions. Western Java, for example, includes the provinces West Java, Jakarta and Banten, all three located on the western side of the island of Java

TABLE A.2
ESTIMATION SAMPLE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cohort	Grade level observed at baseline	Grade level observed at Y2	Grade level observed at Y3	Used in ITT estimation on the Y2 sample	Used in ITT estimation on the Y3 sample	Baseline values available at Y2	Baseline values available at Y3	School average baseline values available at Y2	School average baseline values available at Y3
P1			grade 1	.	1	.	0	.	1
P2		grade 1	grade 2	1	1	0	0	1	1
P3		grade 2	grade 3	1	1	0	0	1	1
P4	grade 2	grade 3	grade 4	1	1	1	1	1	1
P5	grade 3	grade 4	grade 5	1	1	1	1	1	1
P6	grade 4	grade 5	grade 6	1	1	1	1	1	1
P7	grade 5	grade 6		1	.	1	.	1	.
P8	grade 6		
S1			grade 7	.	1	.	0	.	1
S2		grade 7	grade 8	1	1	0	0	1	1
S3		grade 8	grade 9	1	1	0	0	1	1
S4	grade 8	grade 9		1	.	1	.	1	.
S5	grade 9		

Notes: “1”: yes, “0”: no, “.”: Does Not Apply. P1-P8 are primary school cohorts and the S1-S5 are secondary school cohorts in our data. The table shows, by cohort, in which periods we observe them in which grade (columns (1)-(3)), in which analysis we use their test score data (columns (4)-(5)), and whether baseline test scores are available when we observe them in Y2 and Y3 respectively (columns (6)-(9)).

TABLE A.3
FIRST STAGE PROCESS - TEACHERS WHO WERE INELIGIBLE AT BASELINE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Y0 (baseline)			Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Started or completed the certification process	0.02 [0.14]	0.00 [0.00]	0.01** (0.01)	0.07 [0.26]	0.04 [0.21]	0.02 (0.02)	0.22 [0.42]	0.16 [0.37]	0.09** (0.04)
Completed the certification process	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a</i>	0.03 [0.17]	0.00 [0.07]	0.02* (0.01)	0.10 [0.30]	0.06 [0.23]	0.05** (0.02)
Certified and paid the certification allowance	0.00 [0.00]	0.00 [0.00]	0.00 <i>d.n.a</i>	0.02 [0.14]	0.00 [0.00]	0.02** (0.01)	0.03 [0.18]	0.01 [0.11]	0.03** (0.01)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table compares average values between treatment and control groups for teachers who were teaching in sample schools but were not eligible for certification at baseline, according to the criteria that applied at the time. Estimates are based on a model that includes district-triplet fixed effects, which are the strata used for randomization. Within-group standard deviations are reported in brackets in columns (1), (2), (4), (5), (7) and (8). School-level clustered standard errors of the estimated mean differences between treatment and control are reported in parentheses in columns (3), (6) and (9).

TABLE A.4
TESTING FOR DIFFERENTIAL ATTRITION

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: All students					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction staying in the sample	0.87 [0.33]	0.87 [0.34]	-0.01 (0.01)	0.86 [0.35]	0.87 [0.34]	0.01 (0.01)
Observations	42,980	85,658		18,974	36,255	
	Panel B: Selection on students with Y0 score > mean					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction staying in the sample	0.89 [0.31]	0.90 [0.30]	-0.01 (0.02)	0.87 [0.33]	0.90 [0.31]	0.02* (0.01)
Observations	21,910	43,104		10,907	20,343	
	Panel C: Selection on students with Y0 score ≤ mean					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction staying in the sample	0.85 [0.36]	0.84 [0.36]	0.00 (0.01)	0.83 [0.37]	0.83 [0.38]	0.01 (0.01)
Observations	21,070	42,554		8,067	15,912	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table presents tests on differential attrition. Different cohorts (defined in Table A.2) stay in the sample for multiple rounds of the survey. We have attrition, but attrition rates do not differ systematically between treatment and control. Panel A combines all students, Panel B and C present results based on subsets of the data (above and below mean score on the Y0 baseline test). School-level clustered standard errors reported in parentheses. Within-group standard deviations reported in brackets.

TABLE A.5
TESTING FOR DIFFERENTIAL ENTRY INTO SAMPLE SCHOOLS

	(1)	(2)	(3)	(4)	(5)	(6)
	New cohorts in Y2			New cohorts in Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Average student asset index of new cohorts	0.58 [0.23]	0.57 [0.23]	-0.00 (0.01)	0.60 [0.22]	0.58 [0.22]	0.00 (0.01)
Observations	55,989	110,860		62,378	127,279	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. New cohorts of students enter the sample schools after the intervention. The table reports tests on whether the socioeconomic background of students entering post-treatment are the same between treatment and control. Students were asked 7 simple questions on the availability of different assets. Specifically, they were asked whether there is a TV, a fridge, a hand phone, a bicycle, a motor bike, a car, or a computer at their home. The asset index is the fraction of the available items and may take on values from 0 to 1. Cohort P1 (first graders entering the sample schools for the first time at Y3) are not considered here, as they were not asked the asset questions for budgetary reasons. The table shows that there is no significant differential entry into the sample schools. School-level clustered standard errors reported in parentheses. Within-group standard deviations reported in brackets.

TABLE A.6
INTENT-TO-TREAT EFFECTS ON STUDENT TEST SCORES, BREAKDOWN BY SCHOOL TYPE AND SUBJECT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	All school types					Primary school					Junior secondary school				
Panel A: Student test score data measured at Y2															
	Math	Science	Indon.	English	Pooled	Math	Science	Indon.	English	Pooled	Math	Science	Indon.	English	Pooled
Treatment effect	0.00 (0.03)	-0.01 (0.03)	0.00 (0.02)	-0.03 (0.05)	-0.01 (0.02)	0.04 (0.03)	0.03 (0.03)	0.04 (0.03)	<i>d.n.a.</i>	0.03 (0.02)	-0.05 (0.05)	-0.05 (0.04)	-0.04 (0.03)	-0.03 (0.05)	-0.04 (0.04)
Observations	79,510	79,373	79,510	40,673	279,066	38,700	38,700	38,700		116,100	40,810	40,673	40,810	40,673	162,966
R^2	0.28	0.29	0.28	0.34	0.28	0.25	0.28	0.27		0.25	0.33	0.32	0.30	0.34	0.31
Panel B: Student test score data measured at Y3															
	Math	Science	Indon.	English	Pooled	Math	Science	Indon.	English	Pooled	Math	Science	Indon.	English	Pooled
Treatment effect	0.01 (0.03)	0.02 (0.03)	0.02 (0.02)	-0.03 (0.05)	0.01 (0.03)	0.04 (0.03)	0.04* (0.03)	0.03 (0.03)	<i>d.n.a.</i>	0.03 (0.03)	-0.03 (0.05)	0.01 (0.04)	0.00 (0.04)	-0.03 (0.05)	-0.01 (0.04)
Observations	78,164	78,126	78,164	40,539	274,993	37,587	37,587	37,587		112,761	40,577	40,539	40,577	40,539	162,232
R^2	0.26	0.25	0.21	0.34	0.24	0.23	0.23	0.22		0.22	0.29	0.27	0.21	0.34	0.26

Notes: The table reports intent-to-treat effects on student-level test scores. Estimates are reported separately for Y2 (panel A) and Y3 data (panel B). Test score data are constructed by first standardizing by subject-grade-year (so that mean and variance in the control group are 0 and 1 respectively), then stacked so that the unit of observation is student-subject-year. These test scores are then regressed on a dummy variable indicating a treatment school. The estimated parameter on the treatment dummy is reported in column (1) and (2). The regression model further includes district-triplet fixed effects (the strata used for randomization), baseline standardized student-level test scores, baseline standardized averaged school-level test scores. For observations for which baseline test scores are not observed, the baseline values are set to zero. Two dummy variables indicating observations for which individual baseline test scores or school averaged baseline scores are not observed, are also included in the regression model. Weights are applied to scale the student-subject level data back to the level of the student (only for the pooled results of column (5), (10) and (15)). School-level clustered standard errors reported in parentheses.

TABLE A.7
TEST FOR ENDOGENOUS MATCHING FROM STUDENTS TO TARGET TEACHERS

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: All students					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction of students with a target teacher	0.56 [0.48]	0.55 [0.48]	0.01 (0.02)	0.55 [0.47]	0.53 [0.48]	0.03 (0.02)
Observations	89,997	172,952		74,580	147,834	
	Panel B: Students with asset index > median					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction of students with a target teacher	0.57 [0.48]	0.57 [0.48]	-0.01 (0.02)	0.56 [0.47]	0.55 [0.48]	0.01 (0.02)
Observations	35,935	67,003		31,501	60,121	
	Panel C: Students with asset index ≤ median					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction of students with a target teacher	0.55 [0.49]	0.53 [0.49]	0.02 (0.02)	0.55 [0.47]	0.52 [0.48]	0.04* (0.02)
Observations	54,062	105,949		43,079	87,713	
	Panel D: Students with $T - 1$ test score ≤ mean					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction of students with a target teacher	0.56 [0.49]	0.54 [0.49]	0.01 (0.03)	0.52 [0.48]	0.49 [0.48]	0.03 (0.02)
Observations	20,247	39,699		32,032	63,897	
	Panel E: Students with $T - 1$ test score > mean					
	Y2			Y3		
	Treatment	Control	Difference (F.E)	Treatment	Control	Difference (F.E)
Fraction of students with a target teacher	0.51 [0.48]	0.54 [0.49]	-0.03 (0.03)	0.55 [0.47]	0.54 [0.48]	0.02 (0.02)
Observations	21,306	41,063		31,684	59,287	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Panel A tests whether there is differential matching from students to target teachers. Panels B - E test whether there is differential matching, conditional on socioeconomic characteristics (panel B and C), or on test scores in the previous period (panel D and E). As proxy for socioeconomic characteristics we use an asset index. Students were asked 7 simple questions on the availability of different assets. Specifically, they were asked whether there is a TV, a fridge, a hand phone, a bicycle, a motor bike, a car, or a computer at their home. The asset index is the fraction of the available items and may take on values from 0 to 1. The results suggest that there is no differential matching of target teachers to students. School-level clustered standard errors reported in parentheses. Within-group standard deviations reported in brackets.

TABLE A.8
TESTING TREATMENT EFFECT HETEROGENEITY
(CONTINUED)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel B: Student test score data measured at Y3									
	Fraction target teachers	Total target teachers	Student asset index	Number of students	Size relative to biggest school	Log number of students	Log size relative to biggest school	School-level Y0 score	Teacher's age
Treatment	-0.10 (0.07)	-0.05 (0.04)	-0.06 (0.08)	-0.00 (0.04)	-0.01 (0.04)	0.05 (0.19)	0.01 (0.07)	0.01 (0.02)	-0.03 (0.09)
Covariate	0.00 (0.08)	0.01*** (0.00)	0.20*** (0.02)	0.00*** (0.00)	0.30*** (0.09)	0.10*** (0.03)	0.10*** (0.03)	0.22*** (0.03)	-0.00 (0.00)
Treatment - covariate interaction	0.21 (0.13)	0.01 (0.00)	0.02 (0.02)	0.00 (0.00)	0.06 (0.14)	-0.01 (0.04)	-0.00 (0.04)	0.01 (0.04)	0.00 (0.00)
Observations	274,993	274,993	274,993	274,993	274,993	274,993	274,993	270,578	251,671
R^2	0.24	0.24	0.25	0.24	0.24	0.24	0.24	0.23	0.23

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table examines the heterogeneity in treatment effects. Test score data are constructed by first standardizing by subject-grade-year (so that mean and variance in the control group are 0 and 1 respectively), then stacked so that the unit of observation is student-subject-year. The test score is then regressed on a dummy variable indicating a treatment school, a covariate term (mostly school-level (columns 1-8) and one teacher level covariate (column 9), and the interaction between the treatment indicator and the covariate). The estimated parameters on these three regressors are reported in the table. The rest of the model specification is the same as used for the Table V results presented in the main text. Panel A reports results based on Y2 test score data and panel B reports results based on Y3 test score data. School-level clustered standard errors reported in parentheses. The interaction variables used in analysis are the fraction of target teachers in the school at baseline, the total number of target teachers in the school at baseline, a student asset index constructed as the school mean of 7 different asset availability dummies constructed from baseline data, the number of students per school at baseline, the total number of students per school in proportion to the largest primary school (for primary schools) or secondary school (for secondary schools), the natural log of the number of students per school, the natural log of the relative measure of size, the school averaged student-level test score obtained at baseline, and teacher's age.

APPENDIX A: WHO THINKS HIGHER SALARIES WILL IMPROVE PERFORMANCE OF INCUMBENT TEACHERS?

As Appendix B notes, the standard economic model does not predict that unconditionally higher salaries will improve the performance of incumbent teachers; that is, it does not predict effects on the intensive margin. But there is considerable evidence that stakeholders in education often believe that such intensive-margin effects matter. This appendix gives examples of that view, taken from both developing-country and developed-country contexts. Because of the focus of this study, the majority of the quotations are taken from the education sector, but we also include other quotes to show that the argument is applied more broadly to the civil service.

Appendix B formalizes the intuition implicit in these quotes and derives comparative statics.

A. Teachers in Indonesia

Before the Teacher Law and in the early years of its implementation, it was commonly argued in Indonesia that low pay corrodes the motivation and performance of teachers, even when they have some intrinsic motivation. Teachers' union officials pressed this argument through the media in the year before the Teacher Law was passed¹:

The high absence rate of elementary school teachers is understandable, as they are paid far below their monthly cost of living, said the head of an educators union. Indonesian Teachers Union (PGRI) chairman Mohammad Surya said on Thursday the government lacked appreciation for teachers, who, like other professionals, needed good salaries and a clear status.

. . . A recent study by the SMERU Research Institute for the World Development Report 2004 showed that Indonesia ranked third in the average absence rate of elementary school teachers at 19 percent, following Uganda at 39 percent and India at 25 percent. . . .

. . . *Surya said the government's failure to improve teachers' quality of life would keep the absence rate high, . . .* (Santoso 2004)²

The *Jakarta Post* echoed this argument in another article in the same year entitled "Low Salaries Force Teachers to Moonlight", saying that:

Subur and Wawan are two of millions of teachers in the country, who have to take side jobs to make ends meet. Some say it is noble. However, others blame their side jobs for the increasing absenteeism among teachers in the country. (Suwarni 2004)

¹ Throughout this appendix, we have italicized text for emphasis.

² The minister of education is quoted in the same article as dismissing this argument, but the second sentence of his rebuttal implicitly accepts that teachers' pay influences their performance: "Teachers should realize they need to discipline themselves, as they are carrying out a duty to improve the standard of national education, regardless of their salary. Besides, they also receive allowances." (For clarification, the various financial allowances received by teachers were not generally performance-based, but were top-ups to salaries that were either unconditional or conditional on being in certain locations.)

That article, too, cited the country's high teacher absence rates. In November 2005, just before the Teacher Law was passed, "man in the street" interviews by the newspaper encountered similar arguments. See, for example, quotes from these two respondents, neither of whom was a teacher:

Especially in this day and age when the cost of living is so high, Indonesian teachers simply cannot rely on their salaries to make ends meet. *That explains why many teachers look for side jobs to supplement their income. As a consequence, this hampers teachers' ability to focus on teaching. How can teachers be expected to give their best to students when they don't know where their family's next meal will come from?*

How can we expect to have a better quality education system if teachers are busy looking for additional income outside their schools? While we may have poor facilities or a bad curriculum, as long as we have dedicated and creative teachers we can still have a good education system. Aristotle and Plato only needed to explain subjects in front of their students without having to bother about classrooms or other equipment. So, I believe that with good books and good teachers, we can achieve good quality education. *But to get a good teacher, we must pay them enough to allow them to focus on students and the teaching process.* (Jakarta Post 2005)

A World Bank report notes this argument in Indonesia, soon after the Law was passed:

Supriadi and Hoogenboom (2004) argue that low teachers' salaries have contributed significantly to the decline in status of the profession. Given their low salaries, teachers are often forced to find part-time jobs to supplement their incomes. These part-time jobs are often in low status occupations, such as motorcycle driver, tricycle (becak) driver, street vendor, etc⁶. Also, *the need to seek extra income causes some teachers to neglect their teaching obligations. The high rate of teacher absenteeism demonstrates this phenomenon.* (Jalal et al. 2009)

By the same token, if low pay worsens performance, it makes sense that increasing pay should improve teacher performance. And indeed this is the argument made by many. For example, another newspaper article published as the certification program was phased in asserted that:

the certification remains good news for most, if not all, teachers. They welcome the new policy with expectations that it can indeed improve their welfare. . . . Despite all the issues and the flaws, the teacher certification program remains a hope for many people concerned with education in the country. *Thanks to the promised doubled base payment, the educators will compete to improve their quality and the classic problems of welfare will no longer give them an excuse not to do their best before their students.* (Maulia 2008)

Some of those involved in the planning also had high hopes for the intensive-margin effects of the salary increase:

[Professor] Riyanto, who helped the Education and Culture Ministry design the procedure for the teacher certification program in 2008, admitted that the program's results failed to

meet his expectations. “*We initially assumed that a salary increase would encourage teachers to perform better in schools.* However, it turned out that most certified teachers have done almost nothing to improve their [teaching] skills or competency, making them no different than uncertified ones,” he recently told The Jakarta Post. (Widhiarto 2014)

B. Teachers in the global education discussion

It is not only in Indonesia that we find this argument that low salaries worsen teachers’ motivation and the quality of their teaching, and that conversely higher salaries will improve teaching. At the global level, the same argument appears in numerous recent reports.

The International Labour Organisation’s “Handbook of good human resource practices in the teaching profession” gives two rationales for setting teacher compensation high enough, which we can recognize as encompassing both the extensive-margin and intensive-margin rationales for higher pay:

“All countries need to provide teachers with rewards which meet the two equally important strategic objectives mentioned above: (1) the recruitment, retention and performance needs as defined by the relevant education authority; and (2) the incentives for individuals to become and remain teachers over the full length of a professional career as defined by the education system, as well as *foster dedication to professional responsibilities by enabling teachers and their dependents to live in dignity without taking second jobs.*

Together with a tendency for late or non payment of teachers’ salaries, these are amongst the factors which lead teachers in many countries to take on second jobs, to the detriment of their teaching, morale and well being, or to leave teaching altogether. (ILO 2012)

UNICEF’s report on *Protecting Salaries of Frontline Teachers and Healthcare Workers* argues that:

Studies suggest that *low pay is a key factor behind teacher absenteeism, informal fees, and brain drain, which in turn is a cause for poor child outcomes especially in rural areas.* For example, staff absenteeism in the early 2000s was as high as 35 percent in rural Bangladesh, Lesotho, Ghana, Mozambique and Zambia (UNICEF 2010)

The report proposes as one policy response:

Paying attention to real pay levels to ensure that compensation keeps up with increases in the cost of living in order to minimize the risk of staff absenteeism, brain drain and coping strategies such as informal fees.

Similarly, the UNESCO “Methodological Guide for the Analysis of Teacher Issues” says that:

Status, career, and salary issues all have an impact on the attractiveness of the teaching profession, and therefore on the profile of new teachers, their motivation once hired, as

well as on teacher attrition and the social context. *Absenteeism levels are also influenced both by teacher motivation and by the dispositions through which the teacher has been hired*

UNESCO's *Global Monitoring Report 2009* (UNESCO 2009), although focused on governance issues, also reflects this view of intensive-margin effects:

In Malawi, average teacher salaries are too low to meet basic needs. There, and in many other countries, *teachers often have to supplement their income with a second job, with damaging consequences for the quality of their teaching. . . .*

Poor morale and weak motivation undermine teacher effectiveness. Teacher retention and absenteeism and the quality of teaching are heavily influenced by whether teachers are motivated and their level of job satisfaction. *Evidence suggests many countries face a crisis in teacher morale that is mostly related to poor salaries, working conditions and limited opportunities for professional development* (Bennell and Akyeampong, 2007; DFID and VSO, 2008). . . .

One consequence of low relative pay in Central Asia has been *an increase in the number of teachers seeking to supplement their income through a second job* – a phenomenon that has been extensively documented in most Central Asian countries (Education Support Program, 2006). *This practice can have damaging consequences for the quality of education, with some teachers withholding curriculum to pressure students into private tutoring* (Bray, 2003).

Similarly, the *Global Monitoring Report 2014* (UNESCO 2014) notes that

[w]hen salaries are too low, teachers often need to take on additional work – sometimes including private tuition – *which can reduce their commitment to their regular teaching jobs and lead to absenteeism.*

It is important to stress that these reports by international organizations all advocate multipronged approaches to improving teacher performance. None expresses the belief that raising salaries alone will be enough. Yet in each case, embedded somewhere in the argument is the view that salary increases and decreases have intensive-margin effects on the quality of teaching, as these quotes show.

C. Teachers in other countries

We encounter this argument at the country level in countries other than Indonesia as well.

In the United States, advocates for raising teacher pay commonly cite these intensive-margin effects on teachers' ability to better serve their students. A *San Francisco Chronicle* opinion piece by the co-founder of Teacher Salary Project (TSP) argues that

Teachers want to give their all, but *being financially stressed and moonlighting does not allow them to teach their best.* (Calegari 2015)

To bring this problem alive, the TSP has produced a short documentary about Laney,

a public middle school teacher who works two after-school jobs and spends her nights bartending just so she can afford to stay in the classroom. *Laney fears she won't make enough to pay her bills—and fears even more that she can't give 100 percent to her students because she is so over-worked and exhausted. . . .*

If teachers like Laney were appropriately compensated they would no longer need to work two and three jobs outside of the classroom. Instead of struggling to pay rent, they would be able to fully devote themselves to our nation's children.

"It makes me really upset to think I'm not giving them my best," Laney says in the film. (Teacher Salary Project 2015)

In Peru, a survey and study of teachers finds that they too argue that low pay inhibits performance:

77.5 per cent of the Peruvian teachers interviewed consider that they are "badly" or "very badly" paid. Very often, they need to complement their income with other jobs, which results in less time available for lesson preparation and a focus on teaching. Better salaries could benefit the professionalisation of teaching and would allow teachers to focus more on their careers. (van der Tuin and Verger 2013)

In Ethiopia, teachers surveyed for a study also make this argument:

The issues raised by the research were numerous, but the most significant and most often-mentioned causes of demotivation and low morale were: • inadequate salaries • low respect for and low status of teachers • poor management and leadership. These issues have a significant impact on classroom performance, that is, teachers' ability to deliver good quality education, as well as on levels of teacher retention.

In the case of Cambodia, an NGO study (VSO 2008) cited by some international agencies recommends:

[i]ncreasing the salaries of teachers, school directors and staff of the provincial and district offices of education to a level appropriate to the cost of living and linked to inflation. In every focus group conducted with teachers, the issue of pay emerged as the most powerful de-motivating factor. It is impossible to earn a living on a teacher's salary in Cambodia. This basic need is going to remain the top priority over and above any other aspirations teachers have for the quality of their teaching practice until it is fulfilled.

It goes on to say that

a reasonable salary would make the pressure to earn a living wage less intense, which should have a positive effect on teachers' commitment and practice. (VSO 2008)

And in the case of the Caucasus and Central Asia, UNICEF (2011) finds that

The need to rely on additional income from economic activities outside of school applies specifically to teachers in rural areas. . . . [T]eacher absences during harvesting season

are common and tolerated by the school and the community. For a few weeks of the year, the second job absorbs so much of the teachers' time that they temporarily redefine their professional identities and primarily see themselves as farmers or merchants, and only secondarily as teachers with a part-time teaching job at school.

D. Other civil servants

Similar intensive-margin arguments have been made for other civil servants in Indonesia. One argument is that (consistent with model 3 in Appendix B) because civil servants' salaries have been seen as too low, the Indonesian government has not been able to enforce standards of performance. Commenting on a 2009 Law on Public Services, one scholar writes that

Enforcement of the sanctions contained in the law implicitly takes for granted the power of senior bureaucrats within the state apparatus. This may not accurately reflect the power dynamics within Indonesian public service providers. Examining power relations within the bureaucracy more than three decades ago, one observer noted:

In their routine efforts to gather information, implement decisions, and mobilize employees, superiors were faced with the fact that they often did not have sufficient authority to do these things ... [*Civil servants often argued*] *to outsiders, and to themselves, that because government salaries were so low, superiors did not have a right to demand more than a minimum of obedience from them ... It was recognized at the top, just as it was widely claimed at the bottom, that the government did not have the right to demand more than semi-obedience and half-effort ...* On paper, Indonesian superiors ... had the power to act against transgressors and to require subordinates to work every hour of each day, but it was recognized by everyone that what was written down was not conceded in fact, and that it would be futile to act as if it were. The natural response of employees who suffered cuts in honoraria or incentive money was to work less ... The incapacity, or extreme reluctance, of superiors to punish transgressions occurring at others' or even their own expense permitted a chronic crisis of authority to infect every pore of the government bureaucracy. The result was to work at a snail's pace or, commonly, not to work at all (Conkling 1979: 443–550)

. . . . Weak authority among superiors is likely to persist despite the nominal availability of formal means of punishment, as *civil servants will continue to seek refuge in the rhetoric of insubordination because of low pay.* (Buehler 2011)

Many Indonesians believe that higher salaries should reduce corruption, while also improving performance along other dimensions. Note that some aspects of poor teacher performance – such as excessive absenteeism – straddle the line between underperformance and corruption, and so would be covered by both avenues for improving performance:

“Appropriate compensation will not only *have an impact on staff turnover and on employees' productivity and quality of work, but will also reduce tendencies for civil servants to engage in corrupt practices.*” (Tjiptoherijanto 2008)

A survey of business executives, household, and civil servants in Indonesia published in 2000, several years before the Teacher Law of 2005, showed that this view was widely shared:

“Respondents were asked to rank the main causes of corruption in society from amongst a list of possible reasons. *The results showed a strong consensus among all three groups with more than one-third of households (36%) and business enterprises (37%) attributing the main cause of corruption to low civil servant salaries.* Public officials were even more strongly of this view with over half of them (51%) putting this reason first.

. . . . almost half of the public officials reported receiving unofficial payments. *The argument that low salaries are a cause of corruption assumes that wages are inadequate to meet daily needs,* and thus income has to be supplemented with bribes.” (Partnership for Governance Reform³ 2000)

The report’s authors go on to challenge this assumption, saying “While low salaries as a cause of corruption may be the most widely held belief, the accuracy of this relationship is disputed in the corruption research literature.” Nevertheless, a view that was so prevalent may have contributed to the legislature’s decision to raise teacher salaries.

This argument that higher (unconditional) salaries lead to better civil-servant performance is not unique to Indonesia either. In the case of Cambodia, Korm (2011) argues that:

The prevailing opinion is that the low incomes of public servants have led them to pay too little attention to their official tasks and duties as they have diverted their time and effort to obtaining additional sources of income. They have become involved in corruption and „moonlighting“ in other jobs. Furthermore, it is thought that *public servants have rationalised such behaviour using the argument that low pay justifies their poor performance. Whatever the reason, public service delivery is thought to have suffered significantly*

In Cambodia, civil servants are paid sums that cannot support a decent standard of living. Securing adequate income may then become the first priority in their minds as they need to meet their necessary costs of living. Chew (1997) emphasised *that if civil servants were well paid in relation to the cost of living, their performance would be good because they could concentrate on their work. When they are paid reasonably, they are happy and they perform to the required standard without being constantly concerned about finding more money to support their living. However, where public servants’ pay is very low in relation to the cost of living, their productivity and quality of performance are similarly low.*

As Korm points out, McCourt (2003), in his *Global Human Resource Management* book, summarizes this situation using the old joke: “you pretend to pay us, and we pretend to work.”

³ This partnership included the World Bank, United Nations Development Program, the Asian Development Bank (ADB), and a Governing Board comprising “a number of reform minded individuals including ministers, senior public officials and private entrepreneurs.”

Describing the situation in “many developing and transitional countries”, McCourt says that as a result,

It is difficult for a supervisor to criticize an employee’s poor attendance record when the supervisor knows that it is almost forced on the employee (and supervisors are probably in the same position themselves).

APPENDIX B: THEORETICAL FRAMEWORK

In this section, we develop three classes of models that illustrate possible mechanisms through which an unconditional salary increase on the primary teaching job could increase a teacher's effort on that job. The models are extensions of a standard model where, given that the salary on the primary job is unrelated to performance, the teacher will always exert the minimum level of effort. We extend this model by recognizing that: (1) teaching is a pro-social task from which teachers could derive utility through the learning gains they contribute to; (2) there may be reciprocity and gift exchange in employment contracts (Akerlof 1982; Fehr and Gächter 2000); and (3) communities and administrators may provide non-pecuniary sanctions or rewards based on actual performance relative to expectations (Webb and Valencia 2006, Cotlear 2006).

A. The standard model

We assume that salary (S) on the primary teaching job is not dependent on performance, and that teachers exert at least a minimum amount of effort E_{min}^p (where the p indexes the *primary* job). This minimum level of effort may exceed the effort threshold at which the teacher would be dismissed, because the teacher is assumed to have some level of professionalism or intrinsic motivation (varying across individuals) that sets her minimum level of effort under low-powered incentives (as in Holmstrom and Milgrom 1991). Thus, the E_{min}^p could vary across teachers and should be thought of as the default level of teacher effort when there are no financial incentives.

We also allow for the possibility of secondary jobs. Secondary jobs pay a piece-rate wage (w) for each unit of effort (E^s). Workers are endowed with a fixed amount of effort (E), which they can distribute over effort on the primary job (E^p), on the secondary job (E^s), and on leisure (E^L).¹

$$(B.1) \quad E = E^p + E^s + E^L$$

In this standard setup, a worker derives utility from consumption—which in this static framework (we abstract from savings) is assumed to be equal to total earnings derived from the primary and secondary jobs—and from the effort that is devoted to leisure, E^L . The utility functions U_c and U_L are worker-specific with standard properties. Substituting the effort constraint for E^L , the utility function of the teacher is

$$(B.2) \quad U = U_c(S + wE^s) + U_L(E - E^p - E^s)$$

¹ We prefer to model effort allocation rather than time allocation, since time spent at school does not necessarily imply that effort is put into making children learn. Teachers could be away from their classroom chatting with their colleagues, and in fact fieldwork shows that this phenomenon of “shirking while at work” is quantitatively important in some settings, as in India and Indonesia (Kremer et al 2005, McKenzie et al 2014). We would consider this as effort spent on leisure.

In this setting, the teacher will work at his or her default effort level (E_{min}^p) on the primary job. Extra effort beyond this level provides no additional income and reduces leisure. Further, in this standard model, an unconditional increase in salary will lead to a reduction in hours spent on the secondary job, an increase in leisure, and no change in E^p , consistent with a positive income elasticity of leisure.

We introduce three possible extensions of this standard model, each of which could yield an increase in effort on the primary teaching job in response to an unconditional increase in salary S . As discussed in the text, our policy experiment does not correspond to a precise test of any of these particular extensions. Rather, our aim is to illustrate the theoretical models of worker behavior that predict increased teacher effort in response to an *unconditional* increase in base salary.

B. Pro-social preferences

The first extension is to assume that the teacher also derives utility from her contribution to the human capital (ΔHC) of students in her classroom. In other words, the teacher is assumed to have pro-social preferences (Levitt and List 2007). While not all workers may exhibit such preferences, it is widely believed that teachers partly select into their jobs because they hold such preferences. We model utility from pro-social preferences as

$$(B.3) \quad U_S = U_S(\Delta HC)$$

where HC is the human capital accumulated by children in the classroom of the teacher, and is assumed to depend positively on the effort exerted by the teacher on her primary job:

$$(B.4) \quad \Delta HC = f(E^p)$$

We assume that either one or both of f and U_S have decreasing marginal returns to inputs (and that neither has increasing returns). Hence, the reduced form (\tilde{U}_S) also has positive and decreasing marginal returns with respect to effort on the primary teaching job:

$$(B.5) \quad U = U_c(S + wE^S) + U_L(E - E^p - E^S) + \tilde{U}_S(E^p)$$

With this addition to the standard model, a teacher with pro-social preferences would typically exert more than the default minimum effort (E_{min}^p) on the primary teaching job, because she derives utility from contributing to learning of children.² It is easy to see, then, how such a set-up generates a positive income elasticity of effort on the primary teaching job. Basically, there are now two kinds of effort: "grunt" work (E^S) that yields no intrinsic utility

² The main reason for not incorporating the variation in the extent of pro-social preferences into the variation in the teacher's default effort is that pro-social preferences generate a positive income elasticity of effort on the pro-social task, whereas variation in the teacher's default effort level due to variation in their effort norms will not.

and is only done for income (and the consumption made possible by it), and "meaningful" work (E^p) that also provides some positive utility. In equilibrium, effort is allocated across E^p , E^s , and E^L so that the marginal costs and returns are equalized.

Now, if there is an increase in salary on the primary job (S), the marginal utility through consumption from E^s decreases, which should lead to a reduction in E^s and an increase in E^p and E^L . Thus, E^p and E^L are both normal goods. Less necessity to earn money reduces the need for second jobs, and it results in the teacher shifting attention to the other things in life that she appreciates: leisure and the learning of the children in the classroom.

Note that there are also situations where the model does not predict an increase in effort as a result of an increase in the income from the primary job. If before the salary increase the teacher already devotes the minimum effort to the teaching job, or if she does not have second job, a marginal wage increase will not lead to additional effort on the primary teaching job.

C. Gift exchange

We model the idea of gift exchange (Akerlof 1982, Fehr and Gächter 2000) by assuming that the teacher also includes in her utility function the employer's utility (U_G , where the subscript G refers to Government), and that the weight the teacher attaches to the employer's utility depends on the salary received from the employer. When the teacher receives a gift of additional salary, she reciprocates with additional consideration for the objectives of the employer—in this case the Education Ministry, which derives utility from the learning of children in Indonesian schools. In other words, the teacher becomes more motivated to do her job as a result of the salary increase. The teacher can increase learning of children in the classroom by increasing effort on the primary job. The utility function of the teacher in this case can be represented as:

$$(B.6) \quad U = U_c(S + wE^s) + U_L(E - E^p - E^s) + \beta(S)U_G(E^p)$$

where $\beta(S)$ is the weight the teacher attaches to the objectives of the employer and $U_G(E^p) = U_{GL}(f(E^p))$ is the utility that the employer derives from student learning, which in turn is a function of the effort the teacher devotes to her primary job. $\beta(S)$ increases with S , the unconditional salary paid to the teacher.

This model yields predictions similar to those generated by the pro-social preferences model. As with the teacher's utility in that model, in this gift-exchange model the employer's utility is positively related to effort devoted to the primary job. In this case, however, the weight that the teacher places on the employer's utility increases with the salary paid on the primary job. Because the model's formulation is an extended version of the social preference model, the prediction that effort on the primary job can act as a normal good also holds for the gift exchange model. The effect of a salary increase on leisure, by contrast, is ambiguous in this

case. If, as a result of the increase in salary, the weight that the teacher attaches to her employer's utility increases a lot, then the effort she devotes to leisure could fall.

D. Informal pressure

A third possible mechanism for a positive effort response to an unconditional increase in salary is because communities or head masters will expect better performance from teachers who are paid better. Communities or head teachers may provide teachers with non-pecuniary rewards or sanctions that depend on performance *relative to expectations*. For example, communities or head masters may be willing to accept shirking from teachers if those teachers are seen as underpaid³, while they would be willing to apply sanctions for the same level of effort if teacher salaries were raised.

Recognizing this as a possible way of rewarding teachers makes the unconditional salary increase conditional. Pay for performance is introduced by making the non-pecuniary rewards dependent on salary.

Let the function $g(S)$ denote the effort expected by the community given a teacher's salary (with g increasing in S), and let $E^P - g(S)$ represent the amount by which effort on the primary teaching job exceeds this expectation. The reward the community provides to the teachers in terms of utility of the teacher is modeled by the function U_R (which is assumed to be positive with decreasing marginal utility). The utility function of the teacher can then be represented by:

$$(B.7) \quad U = U_c(S + wE^s) + U_L(T - E^P - E^s) + U_R(E^P - g(S))$$

E. Comparative statics results

We derive the effect of a marginal increase in salary on the primary teaching job (S) on effort devoted to the primary teaching job, the second job and leisure. The results are summarized in table B.1. Depending on the allocation of effort before the salary increase, and the model used, the effect of a marginal increase in salary on the primary teaching job on effort devoted

³ This model of behavior, while not described in any prior formal economic model that we know of, is commonly cited by public-sector employees, supervisors, and even beneficiaries in developing countries. For the health sector in Peru, this behavior was formulated by a hospital manager as: "By 10:30 a.m. most of my doctors have skipped out to their second or third jobs. But, how can I demand [compliance] when I know that on their salary they can't make ends meet?" Cotlear (2006). In the Indonesian civil service generally, it has been argued that "[Civil servants often argued] to outsiders, and to themselves, that because government salaries were so low, superiors did not have a right to demand more than a minimum of obedience from them . . . It was recognized at the top, just as it was widely claimed at the bottom, that the government did not have the right to demand more than semi-obedience and half-effort. . ." Writing on human resource management in developing and transition economies, McCourt (2003) argues that "[w]here low pay persists over a period of years, moonlighting becomes institutionalized, with many employees openly absent for several hours of the working day. It is difficult for a supervisor to criticize an employee's poor attendance record when the supervisor knows that it is almost forced on the employee . . ."

to the primary teaching job is either zero or positive. Effort devoted to second jobs moves in the opposite direction, while the effect on effort devoted to leisure is ambiguous.

TABLE B.1.
COMPARATIVE STATICS RESULTS: HOW A MARGINAL INCREASE IN SALARY AT THE PRIMARY TEACHING JOB AFFECTS EFFORT ALLOCATION

	if in the optimum			
	$E^p = E_{min}^p$ $E^s=0$	$E^p > E_{min}^p$ $E^s=0$	$E^p = E_{min}^p$ $E^s >0$	$E^p > E_{min}^p$ $E^s >0$
	Effect on effort on the primary teaching job			
Pro-social preferences	0	0	0	+
Gift exchange	0	+	0	+
Informal pressure	0	+	0	+
	Effect on effort on the secondary job			
Pro-social preferences	0	0	-	-
Gift exchange	0	0	-	-
Informal pressure	0	0	-	-
	Effect on effort devoted to leisure			
Pro-social preferences	0	0	+	+
Gift exchange	0	-	+	ambiguous
Informal pressure	0	-	+	ambiguous

In the remainder of the appendix we derive the results presented in Table B.1. Note that all of the extensions of the standard model discussed in this appendix can be written in the general form:

$$(B.8) \quad U = U_c(S + wE^s) + U_L(T - E^p - E^s) + V(E^p, S)$$

Table B.2 shows the partial derivatives of the function V , depending on the model that is chosen. In all cases, effort on the primary teaching job contributes positively to utility, but the effect of salary varies depending on the model. The cross-partial derivative is positive, except in the model with pro-social preference, where it does not appear.

TABLE B.2.
PARTIAL DERIVATIVES FOR W

	$\frac{dV}{dE^p}$	$\frac{dV}{dS}$	$\frac{dV}{dE^p dS}$
Pro-social preferences	+	0	0
Gift exchange	+	+	+
Informal pressure	+	-	+

The maximization problem for the teacher can be expressed as:

Maximize

$$(B.9) \quad \max_{E^p, E^s} U = U_c(S + wE^s) + U_L(E - E^p - E^s) + V(E^p, S)$$

subject to

$$E^p > E_{min}^p$$

Let e^p , e^s and e^l denote the values at which the teacher obtains maximum utility before the salary increase. We would like to know how a marginal change in S affects these chosen effort levels of the teacher. In the initial equilibrium, the effort levels could be either at or above the minimum values.

If effort on the primary teaching job is at its minimum level, then this indicates that the marginal utility of effort on the primary job (through W) is less than the marginal disutility of extra effort on the primary job (through the utility from leisure):

$$(B.10) \quad V^{E_p}(e^p, S) < U'_L(E - e^p - e^s) \quad \text{if} \quad e^p = E_{min}^p$$

If the teacher does *not* work in a secondary job in the optimum—that is, if $e^s = 0$ —then this means that the marginal utility (through additional consumption) of providing effort on the second job is less than the marginal disutility of that effort through the utility from leisure:

$$(B.11) \quad U'_c(S)w < U'_L(E - e^p) \quad \text{if} \quad e^s = 0$$

The first order conditions for the interior solution - that is $e^p > E_{min}^p$ and $e^s > 0$ are

$$(B.12) \quad V^{E_p}(e^p, S) = U'_L(E - e^p - e^s) \quad \text{if} \quad e^p > E_{min}^p$$

and

$$(B.13) \quad U'_c(S + we^s)w = U'_L(E - e^p - e^s) \quad \text{if} \quad e^s > 0$$

These conditions yield four possible outcomes for the optimal levels of effort provided to the primary teaching job and to the secondary job. Below, the effects of an increase in S are derived separately for each of these four scenarios:

Scenario 1: $e^p = E_{min}^p$ and $e^s = 0$

Consider the effect of a marginal change in income on the primary teaching job if the teacher has no secondary jobs and exerts the minimum effort on the primary job. Because a marginal

change in S will not affect the inequality conditions (B.10) and (B.11), the effect of a marginal change in income on the effort provided to the teaching job is equal to zero.

Scenario 2: $e^p > E_{min}^p$ and $e^s = 0$

Consider the scenario where in the optimum the teacher has no secondary job, but does work more than the minimum number of hours. In this scenario, conditions (B.11) and (B.12) hold. A marginal change in the salary at the primary teaching job will have no effect on effort in secondary jobs, as (B.11) will still hold. To see the effect on the primary teaching job, differentiate (B.12) with respect to S as follows:

$$(B.14) \quad V^{E^p, E^p}(e^p, S) \frac{de^p}{dS} + V^{E^p, S}(e^p, S) = -U_L''(E - e^p) \frac{de^p}{dS}$$

$$(B.15) \quad \frac{de^p}{dS} = \frac{V^{E^p, S}(e^p, S)}{-U_L''(E - e^p) - V^{E^p, E^p}(e^p, S)}$$

In the social preferences model, this is equal to zero, as the numerator is zero. In the other models, it will always be positive.

Scenario 3: $e^p = E_{min}^p$ and $e^s > 0$

Consider the scenario where in the optimum the teacher has a secondary job, but provides minimum effort to the primary teaching job. In this scenario, conditions (B.10) and (B.13) hold. A marginal change in income on the primary teaching job will have no effect on effort on the primary teaching job, as (B.10) will still hold. To see the effect on the secondary job, differentiate (B.13) with respect to S :

$$(B.16) \quad -U_L''(E - E_{min}^p - e^s) \frac{de^s}{dS} = U_C''(S + we^s)w(1 + w \frac{de^s}{dS})$$

$$(B.17) \quad \frac{de^s}{dS} = \frac{U_C''(S + we^s)w}{-U_L''(E - E_{min}^p - e^s) - U_C''(S + we^s)w^2}$$

In all models, this derivative is negative, meaning that the salary increase reduces effort on the secondary job.

Scenario 4: $e^p > E_{min}^p$ and $e^s > 0$

Now consider the scenario where in the optimum the teacher has a secondary job and also provides more than the minimum effort to the primary teaching job. In this scenario, conditions (B.12) and (B.13) hold. To see the effect on the secondary job, differentiate (B.12) and (B.13) with respect to S

$$(B.18) \quad U_C''(S + we^S)(w + w^2 \frac{de^S}{dS}) = -U_L''(E - e^p - e^S)(\frac{de^p}{dS} + \frac{de^S}{dS}) = \\ V^{E^p, E^p}(e^p, S) \frac{de^p}{dS} + V^{E^p, S}(e^p, S)$$

Solving these two equations with two unknowns for $\frac{de^p}{dS}$ (and omitting for the remainder the arguments of the functions to simplify notation) yields

$$(B.19) \quad \left[-U_L'' - V^{E^p, E^p} - \frac{U_L'' V^{E^p, E^p}}{U_C'' w^2} \right] \frac{de^p}{dS} = V^{E^p, l} + U_L'' \frac{-U_C'' w + V^{E^p, S}}{U_C'' w^2}$$

Inserting the signs of the elements of the equation above reveals the sign of the partial derivative of effort on the primary job with respect to salary.

$$\left[-(-) - (-) \quad - \frac{+}{-} \right] \frac{de^p}{dS} = (+/0) + (-) \frac{-(-) + (+/0)}{-} \\ [+] \frac{de^p}{dS} = +$$

Under this scenario, for all models, raising the teacher's salary increases the effort that she exerts on the primary job. To see the effect of a marginal wage increase on effort devoted to secondary jobs, note that the first equality of (B.18) can be rewritten as

$$(B.20) \quad (U_C'' w^2 + U_L'') \frac{de^S}{dS} = -U_C'' w - U_L'' \frac{de^p}{dS}$$

Noting that $\frac{de^p}{dS} > 0$, it follows that $\frac{de^S}{dS} < 0$; in other words, effort on secondary jobs will fall. Finally, to see the effort devoted to leisure, note that the first equation of (B.18) can also be written as

$$(B.21) \quad V^{E^p, E^p} \frac{de^p}{dS} + V^{E^p, S} = U_L'' \left(\frac{de^l}{dS} \right)$$

It follows that when $V^{E^p, S} = 0$, as in the social preference model, leisure will increase. In the other models, the effect on leisure is ambiguous.

**APPENDIX C: TRANSLATION OF LETTER SENT FROM THE INDONESIAN NATIONAL MINISTRY OF
EDUCATION TO THE DISTRICT EDUCATION OFFICES (COPIED TO HEAD TEACHERS OF TREATMENT
SCHOOLS WITH A LIST OF ELIGIBLE TEACHERS)**

To the head of the district education office:

Following up on our letter dated May 18, 2009 regarding the Evaluation Study of the Teacher Certification Implementation, we would like to inform you that we have already received the schools and teachers data from the schools that have been assigned as the sample for the evaluation study of teacher certification. Based on these data, we have conducted data analysis and grouped the teachers into three groups as follows:

1. Certification participants under the additional quota from year 2009
These are teachers that already fulfilled the requirements for certification in 2009, and have been included in the additional quota from year 2009.
2. Candidates of certification participants under the quota from years 2010-2014
These are teachers that have not yet fulfilled the requirements as certification participants from year 2009, and who will be included as certification participants as soon they fulfill the requirements (eligible starting in year 2010 and ending in year 2014).
3. Teachers who were certified in years 2007-2009 and participants from year 2009 quota
This group consists of teachers who already passed the certification in years 2007-2008 and teachers that were already included as certified teachers in the 2009 quota.

The list of names of the teachers in each of the above groups is provided in Appendix I. In accordance with this grouping, we would respectfully request that you pay attention to and conduct several things as follows:

1. Teachers who are included in Group 1 will be assigned as the participants of the additional teacher certification quota through the Head of the District Education Decree, and will be asked to arrange their portfolio based on Book 3 of the Portfolio Development Guideline. Assignment of teacher certification participants' numbers will be in accordance with the regulation regarding the provision of participation numbers and sequence numbers (digits 11-14), which is the subsequent numbering of the participants from quota from year 2009. The portfolio of these teachers should be sent to the university in accordance with their respective area, by July 15, 2009 at the latest. The process of registering the participants is the same as with the quota from year 2009, and teachers are obligated to fill in format A1 and send it to the nearest Education Quality Assurance Agencies by July 10, 2009 at the latest.
2. Teachers who are in Group 2 will receive priority access and can be immediately appointed to join the certification process of the current year when the candidates have fulfilled the requirements. This regulation is only applicable for years 2010-2014.

3. Teachers who are in Group 3 should be attempted not to be transferred to another school during the implementation of this study.
4. Check the accuracy of teacher data sent by the principals. If there are discrepancies between the actual data and the sent data, please report this to us as soon as possible and send the revised teacher data at latest by July 15, 2009.
5. If we found teacher data that may be invalid from a certain school, for example the student teacher ratio is too small, we will clarify to the respective school, and the evaluation of the portfolios of all teachers from this schools will be put on hold until the clarification process has been completed.

Should you need further information you can contact the Teacher Profession Directorate at the Director General of Quality Improvements of Education and Education Personnel at the following phone number 021-5774122 (through Santi and Ubaidah) or online consultation through website www.sertifikasiguru.org during office hours.

For your attention and cooperation, we thank you.

Signed, Director General of Quality Improvements of Education and Education Personnel.

Copies sent to:

- (1) University Rectors responsible for teacher certification
- (2) Secretary, Director General of Quality Improvements of Education and Education Personnel
- (3) Directors, Director General of Quality Improvements of Education and Education Personnel
- (4) Executive Secretary, Teacher Certification Consortium
- (5) Head of Education Quality Assurance Agencies, as applicable
- (6) Principals of applicable schools

APPENDIX D: CALCULATING THE PRESENT VALUE OF A PAY INCREASE THROUGH INTENSIVE- AND EXTENSIVE-MARGIN EFFECTS

Higher pay for teachers may improve teaching quality through either the intensive margin (increasing the motivation of incumbent teachers) or the extensive margin (attracting new, higher-quality teachers into the profession), or both.

To calculate the present value of a policy of across-the-board pay increases, suppose that E_i and E_e are intensive- and extensive-margin effects on student learning respectively (measured in σ /year). These effects can be seen as the steady state annual effects (i.e. a flow effect) on student learning from the policy, and we assume that both effects are constant through time.

We assume that a career in teaching lasts 40 years, meaning that it takes 40 years after the policy change for all incumbent teachers to be replaced by more effective new teachers. (Note that quit rates from government jobs are very low and so staff turnover will only happen with retirement of older cohorts.) Under this assumption, in the first year after the policy of increasing teacher pay, students benefit only through the intensive-margin effects E_i . In the second year, students continue to benefit from the intensive-margin effect, while also getting $1/40^{\text{th}}$ of the extensive margin effect (since $1/40^{\text{th}}$ of the initial stock of teachers is replaced by higher-quality new teachers). This composite effect $E_i + \frac{1}{40}E_e$ is discounted at rate r , where $\delta = \frac{1}{1+r}$.

The formula for the present value of the pay increase is then:

$$R_0 = E_i + \delta \left(E_i + \frac{1}{40} E_e \right) + \delta^2 \left(E_i + \frac{2}{40} E_e \right) + \dots + \delta^{40} \left(E_i + \frac{40}{40} E_e \right) + \delta^{41} \left(E_i + \frac{40}{40} E_e \right) + \dots$$

R_0 can be rewritten as:

$$\begin{aligned} R_0 = E_i + \delta \left(E_i + \frac{1}{40} E_e \right) + \delta^2 \left(E_i + \frac{2}{40} E_e \right) + \dots + \delta^{40} \left(E_i + \frac{40}{40} E_e \right) + \delta^{41} \left(E_i + \frac{41}{40} E_e \right) \\ + \delta^{42} \left(E_i + \frac{42}{40} E_e \right) + \dots - \left[\delta^{41} \left(\frac{1}{40} E_e \right) + \delta^{42} \left(\frac{2}{40} E_e \right) + \dots \right] = \end{aligned}$$

R_0 can then be split into three components:

$$R_0^i = E_i + \delta E_i + \delta^2 E_i + \dots = E_i [1 + \delta + (\delta)^2 + \dots]$$

$$R_0^e = \delta \frac{1}{40} E_e + \delta^2 \frac{2}{40} E_e + \dots = \frac{E_e}{40} [\delta + \delta^2 2 + \dots]$$

$$R_0^{e(\text{adjustment})} = -\delta^{40} \left[\delta \frac{1}{40} E_e + \delta^2 \frac{2}{40} E_e + \dots \right] = -\frac{E_e}{40} \delta^{40} [\delta + \delta^2 2 + \dots]$$

There are two kinds of infinite sums in the above. Both can be rewritten. Using derivatives, we see that:

$$X(\delta) = 1 + \delta + \delta^2 + \dots = \frac{1}{1 - \delta}$$

$$X'(\delta) = 1 + 2\delta + 3\delta^2 + 4\delta^3 \dots = \frac{1}{(1 - \delta)^2}$$

$$\delta X'(\delta) = \delta + 2\delta^2 + 3\delta^3 + 4\delta^4 \dots = \frac{\delta}{(1-\delta)^2}$$

Based on these relationships we can simplify R_0^i , R_0^e and $R_0^{e(\text{adjustment})}$ as follows:

$$R_0^i = E_i \left[\frac{1}{1-\delta} \right]$$

$$R_0^e = \frac{E_e}{40} \left[\frac{\delta}{(1-\delta)^2} \right]$$

$$R_0^{e(\text{adjustment})} = -\frac{E_e}{40} \delta^{40} \left[\frac{\delta}{(1-\delta)^2} \right]$$

Consequently:

$$R_0 = E_i \left[\frac{1}{1-\delta} \right] + \frac{E_e}{40} \left[\frac{\delta}{(1-\delta)^2} \right] - \frac{E_e}{40} \delta^{40} \left[\frac{\delta}{(1-\delta)^2} \right] = E_i \left[\frac{1}{1-\delta} \right] + E_e \frac{1-\delta^{40}}{40} \left[\frac{\delta}{(1-\delta)^2} \right]$$

So, the present value is a weighted sum of the intensive- and extensive-margin effects E_i and E_e .

Indonesia currently pays about 7 percent interest on its government bonds, which is a good default rate for public expenditure ($r = 0.07$). Thus, $\delta = \frac{1}{1+0.07}$.

By substituting this value in, we find that:

$$R_0(r = 0.07) \approx (E_i \times 15) + (E_e \times 5)$$

These figures show that the intensive-margin effect matters considerably more than the extensive margin effect in present value calculations (because the extensive margin gains accrue further in the future and have to be discounted more as a result).

In the context we study, the intensive-margin effects are around 3 times more important than the extensive-margin effects in determining the present value of the policy change (at $r = 0.07$).

It is easy to see that if the annual cost of the pay raise is C , then the present value of the discounted stream of costs will be $(C \times 15)$, since C is analogous to E_i in the calculations above.

If the pay increases are phased in over 10 years (as was done), this estimate will be a little lower ($C \times 11.5$), but in this case, the extensive- margin effects would also be lower. Since the goal of this exercise is to highlight the implications of immediate costs paid on the intensive margin and the delayed benefits from the extensive margin, we focus on the case where the pay increase is implemented immediately and assume that all new cohorts reflect the extensive-margin effects. Note also that the cost estimates are conservative, because future pay increases are usually a percentage of base pay and we assume that there are no increases in real teacher salaries. If real wages grow at 2% or 3% per year, the present value of total costs would be $(C \times 21)$ and $(C \times 27)$ respectively.