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

de Ree, J.; Muralidharan, K.; Pradhan, M.; Rogers, H.

DOI
10.1093/qje/qjx040

Publication date
2018

Document Version
Final published version

Published in
The Quarterly Journal of Economics

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

Joppe de Ree
Karthik Muralidharan
Menno Pradhan
Halsey Rogers

How does a large unconditional increase in salary affect the performance of incumbent employees in the public sector? We present experimental evidence on this question in the context of a policy change in Indonesia that led to a permanent doubling of teacher base salaries. Using a large-scale randomized experiment across a representative sample of Indonesian schools that accelerated this pay increase for teachers in treated schools, we find 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, and we can rule out even modest positive impacts on test scores. Our 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. *JEL Codes*: J31, J45, I21, C93, O15.

*We are especially grateful to Gordon Dahl and Lawrence Katz (the editor) for extensive comments on multiple drafts of this article. 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 *puslitjak*), 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 benefited from the 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 article are entirely those of the authors. They do not necessarily

The Quarterly Journal of Economics (2018), 993–1039. doi:10.1093/qje/qjx040.
Advance Access publication on November 13, 2017.

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, e.g., Govt. of India 2008, 2015). 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.

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).

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.

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 seventh Pay Commission in India raised government expenditure by 0.65% of GDP and required forgoing or deferring capital investments to meet fiscal deficit targets (Govt. of India 2015; Sabnavis and Shah 2015).

In this article, 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 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. The study was conducted over a three-year period, in an almost 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

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).

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 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).

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 important, we find no difference in student test scores in language, mathematics, or science across treatment and control schools. The point estimates are close to 0 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 nonparametric 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.

3. Roughly 20% of teachers in both treatment and control 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.

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 nonincentivized prosocial 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, Falk 2007). 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 perspective, we are more interested in whether such an expensive policy (which costs over 5% of the national budget) improved the effort of incumbent teachers and learning outcomes of their students through any combination of the posited mechanisms. 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 article is Mas (2006), which finds that police performance in New Jersey deteriorated when arbitrators awarded

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 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 nonexistent (Kube, Maréchal, and Puppe 2013, Esteves-Sorenson forthcomimg).

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 Kőszegi 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 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 United States (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 nonwage 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 2011a; 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).

Our results do not imply that salary increases for public employees would have no positive effects 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; Donohue, Heckman, and Todd 2002).

Instead, 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 although the policy improved 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 policy making in a public-sector 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]

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).

6. In contrast, Henry Ford's famous "five-dollar workday" led to a similar doubling in wages and 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

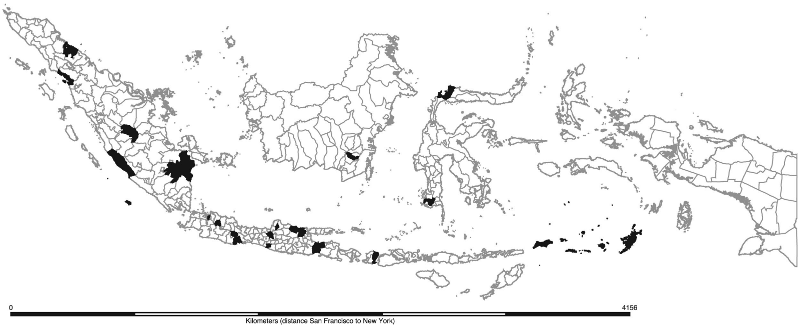

FIGURE I

Map of Indonesia with 20 Selected Districts

The rest of this article 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 appendixes 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 more than 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 2000–2002, when the new Indonesian constitution committed the government

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.

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 on 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 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-standard 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-standard certification exams. Thus, by the time the final law and regulations were negotiated through the political and policy-making 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. In practice, the certification process yielded a doubling of base pay with only a modest hurdle to be surmounted.[8]

Prereform 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

7. Note that the professional allowance was 100% of base pay, rather than of total precertification 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. Furthermore, 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 to be submitted with the certification process (World Bank 2010).

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 prereform 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 increase was not conditional on teachers' subsequent effort or effectiveness, but 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 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 partly 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). Although 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 "low salaries are likely to damage (teacher) morale" 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 "low 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]

Online 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 Online Appendix B, we formalize the economic arguments implied in these quotes from practitioners, presenting simple theoretical sketches of mechanisms by which teacher effort may increase in response to an unconditional pay increase and deriving comparative statics. These include: (i) reciprocity and gift exchange in employment contracts; (ii) a model in which effort on prosocial tasks like teaching is a normal good with a positive income elasticity; and (iii) 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]

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 a 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

9. The argument that higher teacher salaries can improve motivation and performance appears in the U.S. 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.

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 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. 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 already. 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 nearly representative sample of 360 schools across 20 districts of Indonesia to make up 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]

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

11. The district sampling for the two parallel sets of randomized evaluations was conducted using the same procedures, 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 nearly 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).

13. We dropped the strata comprising schools with very large and very small numbers of teachers. If schools were too large, it would not have been feasible to test all the students in the school during the time 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

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 begin the process (a translated copy of this letter appears in Online 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

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 ineligible teachers in treatment schools once they met the eligibility criteria (point 2 in the letter). 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.

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 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. Furthermore, 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 nontarget 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 preannounced 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 article. 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 were 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).

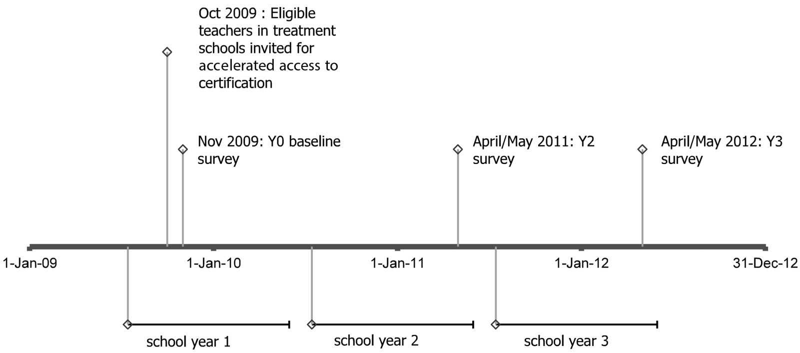

FIGURE II

Project Timeline

We carried out three waves of data collection, during which we interviewed head teachers, teachers, and students; we conducted independent tests of both teacher knowledge and student learning outcomes. The first wave was a baseline collected in November 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 two years of the project (Y2), and the third wave was collected in April–May 2012, at the end of three 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

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.

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 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 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

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.

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

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

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. 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 are 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.

the absence of the experiment). We observe small differences in a few other teacher characteristics that are attributable to random sampling variation. The magnitude of these differences is small, especially when compared with 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]

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

TABLE II
BALANCE TESTS ON TEACHER-LEVEL VARIABLES

	All teachers			Target teachers only		
	Treatment (1)	Control (2)	Difference (F.E) (3)	Treatment (4)	Control (5)	Difference (F.E.) (6)
Raw (fraction correct) test score	0.56	0.56	0.00	0.55	0.56	−0.01
	[0.16]	[0.16]	(0.01)	[0.17]	[0.17]	(0.01)
Eligible but not certified at baseline (i.e., target)	0.56	0.57	−0.01	1.00	1.00	−0.00
	[0.50]	[0.50]	(0.02)	[0.00]	[0.00]	(*d.n.a.*)
Already certified at baseline	0.19	0.18	0.02	0.00	0.00	0.00
	[0.39]	[0.38]	(0.01)	[0.00]	[0.00]	(*d.n.a.*)
Not eligible for certification at baseline	0.25	0.25	−0.00	0.00	0.00	0.00
	[0.43]	[0.43]	(0.01)	[0.00]	[0.00]	(*d.n.a.*)
Bachelor's degree	0.62	0.59	0.04***	0.69	0.65	0.06***
	[0.49]	[0.49]	(0.02)	[0.46]	[0.48]	(0.02)
High rank (rank IV) in civil service	0.41	0.44	−0.03	0.48	0.51	−0.04**
	[0.49]	[0.50]	(0.02)	[0.50]	[0.50]	(0.02)
Certified and paid the certification allowance	0.11	0.12	−0.01	0.00	0.00	0.00
	[0.32]	[0.33]	(0.01)	[0.00]	[0.00]	(*d.n.a.*)
Base pay (in mil. IDR)	1.87	1.92	−0.05	2.02	2.07	−0.07**
	[0.83]	[0.80]	(0.03)	[0.73]	[0.69]	(0.03)
Other allowances (in mil. IDR)	0.53	0.54	−0.02	0.55	0.59	−0.04***
	[0.34]	[0.33]	(0.01)	[0.31]	[0.31]	(0.01)

TABLE II
CONTINUED

	All teachers			Target teachers only		
	Treatment (1)	Control (2)	Difference (F.E) (3)	Treatment (4)	Control (5)	Difference (F.E.) (6)
Certification pay (in mil. IDR)	0.21	0.22	−0.01	0.00	0.00	0.00
	[0.59]	[0.60]	(0.02)	[0.00]	[0.00]	(*d.n.a.*)
Second job	0.34	0.34	−0.00	0.33	0.35	−0.02
	[0.47]	[0.47]	(0.02)	[0.47]	[0.48]	(0.02)
Hours worked on second job	3.50	3.40	0.12	3.18	3.40	−0.19
(last week)	[8.04]	[7.69]	(0.29)	[6.99]	[7.48]	(0.34)
Started or completed the	0.61	0.29	0.33***	0.73	0.18	0.54***
certification process	[0.49]	[0.45]	(0.02)	[0.45]	[0.39]	(0.03)

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. 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). *d.n.a.* refers to cases for which standard errors do not apply.

We also test for differential attrition and entry of students over the period of the study. Online Appendix 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 (Online Appendix Table A.4, Panel A) and 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 (Online Appendix 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 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

fraction of certification-eligible teachers across treatment and control schools (56% versus 57%; columns (1) and (2)). There are small differences in precertification 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).

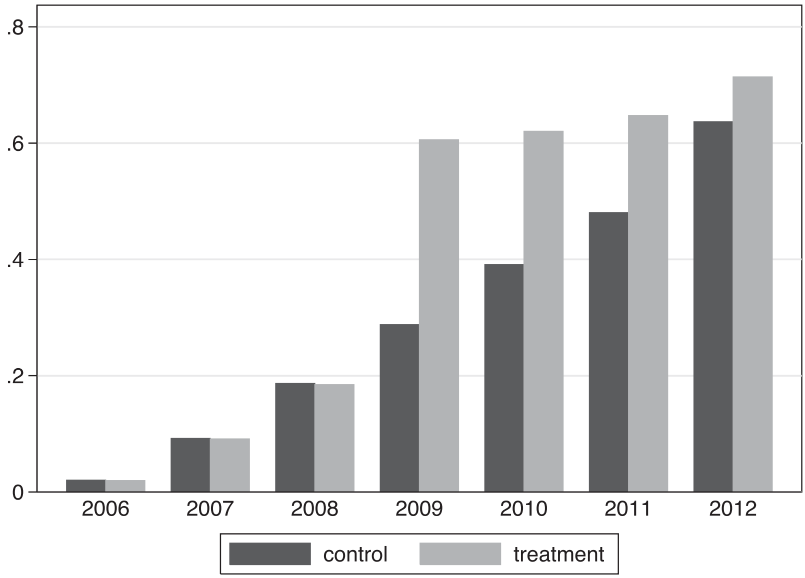

FIGURE III

Fraction of Teachers Admitted to the Certification Process at or before the Indicated Year

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 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.

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 II and IV, 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

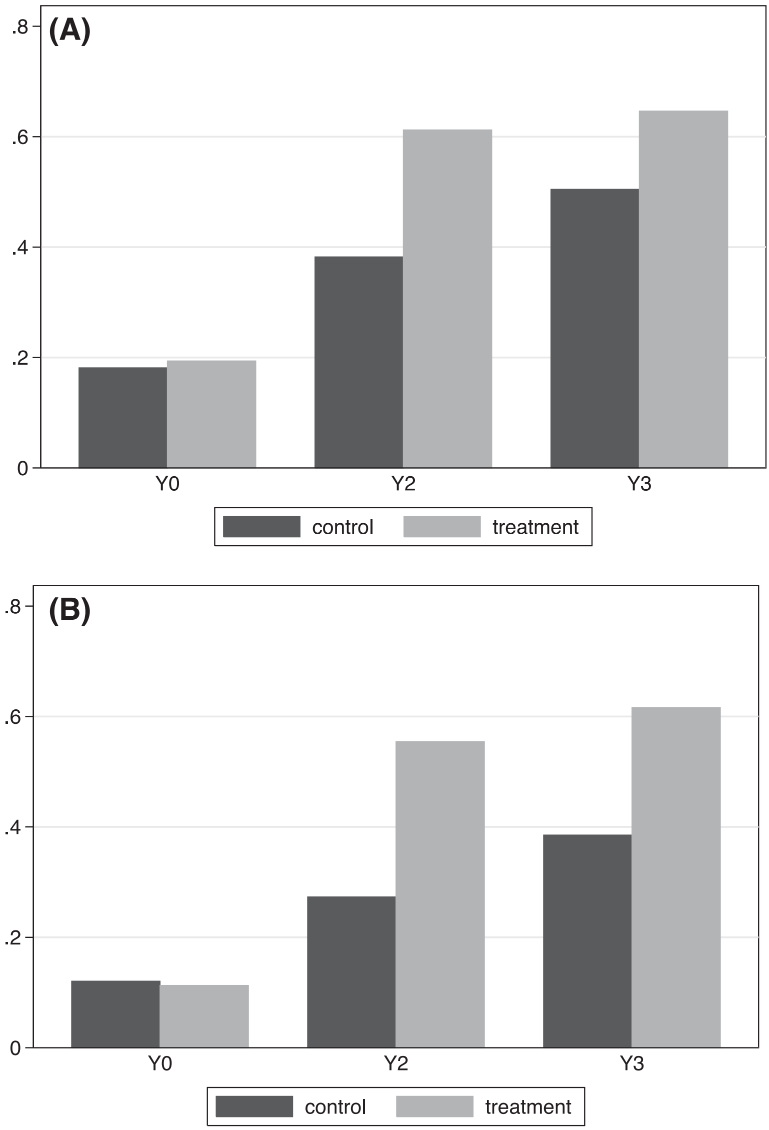

FIGURE IV

Completing the Certification Process (A) and Being Paid the Certification Allowance (B).

Panel A presents the fraction of teachers who completed the certification process. Panel B presents the fraction of teachers who completed the certification process and were paid the certification allowance.

TABLE III

FIRST STAGE PROCESS - TEACHER LEVEL

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

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. 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). *d.n.a.* refers to cases for which standard errors do not apply.

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 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. 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).

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.

Finally, we present the corresponding figures for the nontarget teachers, who were not eligible for certification at the start of the experiment in Online Appendix 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 Online Appendix C). However, as Online Appendix Table A.3 shows, 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 increase). We report both simple differences (with stratum fixed effects) and differences-in-differences 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

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 differences-in-differences 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 $\frac{0.64}{0.57}$; this is the 112% figure reported in the text.

TABLE IV

TEACHER-LEVEL IMPACT

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

TABLE IV
(CONTINUED)

	All teachers						Target teachers only					
	Y2			Y3			Y2			Y3		
	Control mean (1)	ITT (simple diff.) (2)	ITT (diff. in diff.) (3)	Control mean (4)	ITT (simple diff.) (5)	ITT (diff. in diff.) (6)	Control mean (7)	ITT (simple diff.) (8)	ITT (diff. in diff.) (9)	Control mean (10)	ITT (simple diff.) (11)	ITT (diff. in diff.) (12)
Total pay (in	3.41	0.44***	0.66***	4.29	0.43***	0.64***	3.62	0.92***	1.11***	4.48	0.89***	1.03***
mil. IDR)	[1.97]	(0.08)	(0.05)	[1.95]	(0.09)	(0.07)	[1.58]	(0.08)	(0.06)	[1.70]	(0.10)	(0.08)
Financial	0.50	−0.09***		0.56	−0.09***		0.48	−0.13***		0.51	−0.16***	
problems	[0.50]	(0.02)		[0.50]	(0.02)		[0.50]	(0.02)		[0.50]	(0.03)	
Satisfied with	0.60	0.09***		0.60	0.07***		0.60	0.17***		0.65	0.13***	
total income	[0.49]	(0.02)		[0.49]	(0.02)		[0.49]	(0.02)		[0.48]	(0.02)	
Absent from	0.14	−0.00	−0.02	0.13	0.01	−0.00	0.12	−0.03*	−0.04**	0.10	0.00	−0.01
school at least	[0.34]	(0.01)	(0.01)	[0.33]	(0.01)	(0.01)	[0.32]	(0.02)	(0.02)	[0.31]	(0.02)	(0.02)
once last week												
Baseline controls		no	yes		no	yes		no	yes		no	yes

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. The table reports intent-to-treat effects at the teacher level. Columns (1), (4), (7), and (10) report average 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.

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 precertification 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]

Because 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 two to three years of base pay. Furthermore, 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 satis-

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 $\frac{0.57\text{M}}{0.27}$, 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 precertification pay $\left(\frac{2.11\text{M}}{2.85\text{M}}\right)$. 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 $\left(\frac{0.38\text{M}}{0.18}\right)$ in the control group, which is also 2.11M IDR. But since other allowances for senior civil service teachers were higher, the total precertification pay for the "target" teachers was 3.25M IDR. Thus, target teachers received a 65% increase $\left(\frac{2.11\text{M}}{3.25\text{M}}\right)$ in their total pay on 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).

faction, reduced financial stress, and reduced outside jobs should be interpreted as the result of an increase in two to three 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 Online Appendixes 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

1. ITT Estimates. Since the randomization was conducted at the school level, we first present school-level ITT estimates. These estimates quantify how student learning in a school responds to a

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.

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:

$$T_{ijks}\ (Y_n) = \beta_0\ + \beta_{1\mathrm{a}} \cdot \overline{T_{ijks}}\,(Y_0) + \beta_{1\mathrm{b}} \cdot T_{ijks}\,(Y_0) + \beta_2 \cdot Treatment_k$$
$$+\ \beta_{Z_{ST}} \cdot Z_{ST} + \varepsilon_{ijks}. \tag{1}$$

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 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. Online Appendix Table A.6 presents results individually for each subject, by school type (primary and junior secondary), and at the end of Y2 and Y3 (Panels A and B); the

25. As we show in Online Appendix Table A.2, some of the cohorts included in our analysis did not have a baseline test. We set the normalized baseline score to 0 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.

TABLE V

INTENT-TO-TREAT EFFECTS ON STUDENT TEST SCORES

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

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. 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 columns (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 0. 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 are reported in parentheses.

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 Y3 (Panels B and D). We see that the treatment effects are not only zero on average but cannot be statistically distinguished from zero at any part of the test score distribution. In Panels A and C, 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 nontarget teachers (especially those who were not eligible for certification), who may have reduced effort in response to the perceived unfairness of not receiving the

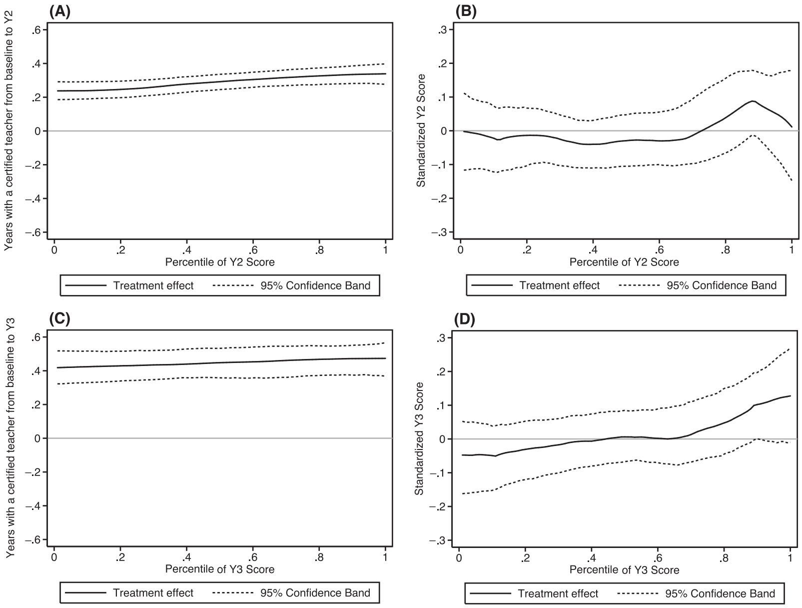

FIGURE V

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

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 0 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.

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 nontarget 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

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 ineligible teachers to feel such resentment. But we still test for this possibility.

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

	Y2 (1)	Y3 (2)
Conditional treatment effect for students with:		
Target teacher in Y_2	−0.004	
	(0.026)	
Nontarget teacher in Y_2	−0.001	
	(0.030)	
Target in Y_2 and Y_3		−0.023
		(0.033)
Target in Y_2 and nontarget in Y_3		0.055
		(0.041)
Nontarget in Y_2 and target in Y_3		0.020
		(0.039)
Nontarget in Y_2 and Y_3		0.029
		(0.036)
No match between student and teacher	−0.092	−0.045
	(0.056)	(0.049)
H_0: causal parameters are the same (p-value)	.91	.26
H_0: causal parameters are the same and equal to zero (p-value)	.99	.37
Total student-subject observations	279,066	274,993

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. 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 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 Y1 outcome data (column (1)) we present causal parameters for students with target teachers in Y2, and for students who do not have target teachers in Y2. For Y3 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 Y2. For column (2) results, about 12.2% of the student-level observations were not matched to teachers in both Y2 and Y3. 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.

nontarget 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 nontarget teachers in treatment schools.[27]

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").

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 – nontarget; nontarget – target; and nontarget – nontarget) and again find no significant difference 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 nontarget teacher in both Y2 and Y3—we still find no evidence that the former did better in treated schools (Table VI, column (2)).

2. Instrumental Variable (IV) Estimates. The ITT estimates 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:

$$\begin{aligned} T_{ijks}\ (Y_2) = \beta_0\ + \beta_{1\mathrm{a}} \cdot \overline{T_{ijks}}\,(Y_0) + \beta_{1\mathrm{b}} \cdot T_{ijks}\,(Y_0) \\ + \beta_2 \cdot Certified_{ijks}\,(Y_2) + \beta_{Z_{ST}} \cdot Z_{ST} + \varepsilon_{ijks}, \end{aligned} \tag{2a}$$

$$\begin{aligned} T_{ijks}\ (Y_3) = \beta_0\ + \beta_{1\mathrm{a}} \cdot \overline{T_{ijks}}\,(Y_0) + \beta_{1\mathrm{b}} \cdot T_{ijks}\,(Y_0) \\ + \beta_2 \left[Certified_{ijks}\,(Y_3) + \Upsilon \cdot Certified_{ijks}\,(Y_2)\right] \\ + \beta_{Z_{ST}} \cdot Z_{ST} + \varepsilon_{ijks}, \end{aligned} \tag{2b}$$

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 equation (1).

One technical consideration in estimating equation (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 Υ (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 equation (2b) for a range of values of Υ and present the resulting estimates of β_2, along with standard errors, in Table VII. The estimates with $\Upsilon = 0$ correspond to complete decay of any test score gains in a year by the end of the next year, while those with $\Upsilon = 1$ correspond to complete persistence. Based on several prior studies, our preferred estimates assume $\Upsilon = 0.5$.

The main threat to interpreting these estimates as the annual impact of being taught by a certified teacher is the possibility of endogenous reassignment of certified teachers within treatment schools to potentially weaker students. We test for this in Online Appendix 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 (Online Appendix 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 (Online Appendix 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

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 nontarget teachers (as seen in Table VI).

TABLE VII

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

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

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$. The table reports IV estimates where effective number of years with a certified teacher since baseline is instrumented with the dummy variable indicating a treatment school. The variable effective number of 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 Y2 and Y3 samples, as well as selected groups who were more strongly affected by the intervention. Columns (3) and (9) report upper bounds 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.

student affluence, measures of school size, and mean baseline test scores, and find no evidence of any heterogeneous effects (Online Appendix 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 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.

Because 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

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.

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.

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 a 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 a 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 a year through extensive margin effects at a cost of US$ 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 a year (averaged across math and language) at an annual cost of only about US$ 4 per student, including implementation costs (Muralidharan and Sundararaman 2011a).[34]

32. The assumption is not unrealistic in theory because 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 (six years after the reform), nearly 50% of recently recruited teachers (between 24 and 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), 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 9,000 IDR per 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.

34. Incentive treatments cost up to 10,000 rupees 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 US$2 per student. The authors conservatively estimate the

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 Online 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

cost of implementing the program as equal to the costs of the bonuses; including the implementation cost would double the per child cost to US$4 per student, which is the figure we use.

35. The present discounted value of a continuous stream of annual costs C, is equal to $\frac{C}{1-\delta}$, where δ is the discount factor, which is equal to $\frac{1}{1+r}$, where r is the discount rate. Thus, if r is 7%, then $\frac{1}{1-\delta}$ 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 Online Appendix D). Note also that from a public budgeting perspective, 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

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 immediately), while the present discounted value of the stream of benefits is only five 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 benefits 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). Although 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. Furthermore, 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

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.

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 underappreciated 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).

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.

VI. Discussion and Conclusion

This article 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 policy makers hoped 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 significantly 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 toward 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 Online Appendixes 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. Furthermore, 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 health care

37. One policy option that mitigates this problem is to have the higher salaries apply only 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.

workers, and Muralidharan and Sundararaman 2011a 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 2011a). 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.

World Bank
University of California, San Diego,
National Bureau of Economic Research,
Bureau for Research and Economic Analysis of Development,
and Abdul Latif Jameel Poverty Action Lab
University of Amsterdam, Vrije Universiteit Amsterdam,
Tinbergen Institute, and Amsterdam Institute for Global
Health & Development
World Bank

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online. Data and code replicating the tables and figures in this article can be found in De Ree et al. (2017), in the Harvard Dataverse, doi:10.7910/DVN/MTVM5O.

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,” NBER Working Paper 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*, 33 (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,” Policy Research Working Paper WPS 8050, World Bank Group, 2017.
Betts, Julian R., “Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth,” *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,” *Journal of Political Economy*, 100 (1992a), 1–40.
———, “School Quality and Black-White Relative Earnings: A Direct Assessment,” *Quarterly Journal of Economics*, 107 (1992b), 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,” *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.
De Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers, “Replication Data for: ‘Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia’” *Harvard Dataverse*, (2017), doi:10.7910/DVN/MTVM5O
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, John J. III, James J. Heckman, and Petra Todd, “The Schooling of Southern Blacks: The Roles of Legal Activism and Private Philanthropy, 1910–1960,” *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,” NBER Working Paper 14906, 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. (Amsterdam: 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," *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?," NBER Working Paper 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," International Monetary Fund, 2016.
Jayaraman, Rajshri, Debraj Ray, and Francis de Véricourt, "Anatomy of a Contract Change," *American Economic Review*, 106 (2016), 316–358.
Kőszegi, Botond, and Matthew Rabin, "A Model of Reference-Dependent Preferences," *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 (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 (2000), 393–408.
Mas, Alexandre, "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 121 (2006), 783–821.
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 Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119 (2011a), 39–77.
———, "Teacher Opinions on Performance Pay: Evidence from India," *Economics of Education Review*, 30 (2011b), 394–403.
———, "Contract Teachers: Experimental Evidence from India," NBER Working Paper 19440, 2013.
Raff, Daniel M. G., and Lawrence H. Summers, "Did Henry Ford Pay Efficiency Wages?," *Journal of Labor Economics*, 5 (1987), S57–S86.
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, 2014.

UNICEF, *Teachers: A Regional Study on Recruitment, Development and Salaries of Teachers in the Ceecis Region* (Geneva: UNICEF, 2011).

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. (Washington, DC: World Bank, 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).

———, *Teacher certification and beyond: An empirical evaluation of the teacher certification program and education quality improvements in Indonesia* (Washington, DC: World Bank, 2015).