KIAT Guru

Teacher Accountability and Pay-for-Performance Schemes in (Semi-) Urban Indonesia:

What do Education Stakeholders Think?

Marcello Perez-Alvarez, Jan Priebe & Dewi Susanti

January 2020

Australian Government
SUPPORTED BY THE WORLD BANK IBRD • IDA | WORLD BANK GROUP
LOCAL SOLUTIONS TO POVERTY
BaKTI

1818 H Street NW, Washington DC 20433
Telephone: 202-473-1000; Internet: www.worldbank.org

Teacher Accountability and Pay-for-Performance Schemes in (Semi-) Urban Indonesia:

What do Education Stakeholders Think?

Marcello Perez-Alvarez, Jan Priebe & Dewi Susanti[1]

SOCIAL DEVELOPMENT UNIT
WORLD BANK – INDONESIA
JANUARY 2020

1 Authors' affiliations, respectively:
(a) University of Göttingen, Department of Economics, e-mail: marcello.perez@wiwi.uni-goettingen.de
(b) GIGA Institute Hamburg, University of Göttingen, e-mail: jpriebe@uni-goettingen.de
(c) World Bank, e-mail: dsusanti@worldbank.org

Table of Contents

List of Figures

List of Table

Acknowledgements

Financial support for this paper was generously provided by the Government of Australia's Department of Foreign Affairs and Trade (DFAT) through the Supporting 12 Years Quality Education for All (ID-TEMAN) Trust Fund. We would like to thank Dedy Junaedi, Lulus Kusbudiharjo, Anas Sutisna, Mulyana, and all survey enumerators for the data collection; Rajius Idzalika for her support on data analysis; Tazeen Fasih, Susiana Iskandar, Lily Hoo, Audrey Sacks, and Yulia Herawati for technical and analytical advice; Noah Yarrow, Amer Hasan, Samer Al-Samarrai, and Ezequiel Molina for their comments on the earlier version of this paper; Bryan Rochelle for editing; Yohanes Cahyanto Aji for report typesetting and formatting, and Dinda Putri Hapsari for report format proofreading.

Disclaimer: the views expressed in this publication are the authors' alone and are not necessarily the views of the Australian Government or the World Bank.

Abstract

Teacher evaluations are conducted to inform employment decisions and teacher professional development with the ultimate goal to create beneficial student learning environments. The effectiveness and feasibility of teacher evaluations, particularly in high-stakes contexts (hiring, firing, promotion, Pay-for-Performance schemes), crucially depends on the support these evaluations receive from the various education stakeholders involved. While many governments around the world, including the Government of Indonesia, are interested in reforming and expanding their current teacher evaluation systems, often little is known about how principals, teachers, parents and students perceive these evaluations.

This paper uses data from a recent large-scale opinion survey in Indonesia to examine and provide rare insights into the attitudes of key education stakeholders towards teacher performance evaluations. Four key insights are identified. First, many principals and teachers agree with existing evaluation schemes employed in Indonesia, such as the teacher competence test (*Ujian Kompetensi Guru* or UKG) and the teacher performance evaluation (*Penilaian Kinerja Guru* or PKG), and are also open to reforms and the introduction of new schemes. Second, Pay-for-Performance schemes are generally popular among principals and teachers, and preferred over seniority-linked pay systems. Third, teachers in urban areas are more favorable towards Pay-for-Performance schemes than teachers in semi-urban areas. Finally, all stakeholders generally support the concept of principals, teachers and parents fulfilling performance evaluator roles.


Saya Sebagai Anak Perempuan atau

01 Introduction

Net enrolment

in primary schooling

rate above

90%

in secondary schooling

increased from

54%

in 2003

to

76%

in 2015

National budget

20%

of the national budget has been allocated to the education sector since 2009

Education budget

in 2017

52%

was allocated to teacher salaries and allowances, with the TPG taking up

35.2%

of that share

Over the last 15 years, Indonesia has made notable progress and investments in improving both access to, and attainment of, education. Net enrolment in primary schooling has remained high at rates above 90 percent, while net enrolment rates in secondary schooling have increased from 54 percent in 2003 to 76 percent in 2015 (World Bank 2018a). At the same time, the Government of Indonesia (GoI) has made remarkable fiscal efforts to improve the quality and effectiveness of education services and outcomes. As a result of the Law on the National Education System (No. 20/2003), Indonesian public education expenditure has more than doubled during the twenty-first century. Moreover, the mandated 20 percent of the national budget has been allocated to the education sector since 2009 (Chang et al. 2014).

Most GoI fiscal efforts have been dedicated to increasing teacher salaries. In 2005 the GoI passed the Teacher Law, aimed at raising the quality and motivation of the teaching force. A major component of the Teacher Law has been the introduction of a teacher certification process. To be certified, teachers have to pass certain education quality standards in order to obtain a teacher professional allowance (*Tunjangan Profesi Guru*, hereafter TPG) that effectively doubles their base salary (World Bank 2010). As a result, the payment of the TPG has put sizable pressure on the GoI's fiscal budget. In 2017, 52 percent of the total education budget was allocated to teacher salaries and allowances, with the TPG taking up 35.2 percent of that share (Ministry of Finance 2016).[2]

The introduction of TPG, however, has not achieved any recognizable progress in improving student learning outcomes (de Ree et al. 2018). This is despite teachers being more satisfied with their salary and less likely to pursue additional jobs outside of their regular teaching duties following the introduction of TPG (de Ree et al. 2018). Indonesian students continue to rank at the bottom of the learning distribution in the Programme for International Student Assessment (PISA) 2015 study, taking the 66th place among 72 participating countries (OECD 2016).[3,4] Likewise, Indonesian student learning outcomes are particularly weak in rural compared to urban areas, a result that can be partially attributed to both worse school infrastructure and higher teacher absence rates in these areas (ACDP 2014).

To improve service delivery and raise student learning outcomes, de Ree et al. (2018) propose the introduction of strong teacher accountability mechanisms, namely Pay-for-Performance (PfP) schemes. This recommendation follows findings from international literature that suggest PfP schemes can improve service delivery and raise student learning outcomes, particularly in low and middle-income countries (Bruns and Schneider 2016; Jinnai 2016; Evans and Popova 2015; Chang et al. 2014; Holla et al. 2012; Pradhan et al. 2014; Joshi 2013; Kremer, Brannen

2 This is equivalent to USD 5.5 billion according to the authors' own calculations based on various published government expenditure reports.

3 This figure refers to the total number of countries participating, and comprises both entire countries and specific administrative areas of countries such as Hong Kong-China and Macao-China.

4 Many poor developing countries do not participate in the PISA (Programme for International Student Assessment) study. Therefore, Indonesia's student learning results should be interpreted as being low compared to other middle and high income countries participating in the PISA program.

and Glennerster 2013; Muralidharan and Sundararaman 2011a, 2011b; Bruns, Filmer and Patronis 2011; Glewwe, Ilias and Kremer 2010; Murnane and Cohen 1986)at least six systematic reviews or meta-analyses have examined the interventions that improve learning outcomes in low- and middle-income countries. However, these reviews have sometimes reached starkly different conclusions: reviews, in turn, recommend information technology, interventions that provide information about school quality, or even basic infrastructure (such as desks. Similarly, the World Development Report 2018 proposes the use of both pecuniary and non-pecuniary incentives to improve teacher motivation and student learning outcomes (World Bank 2018b).

The introduction of teacher PfP elements is a rather new initiative for Indonesia. However, an ongoing pilot (KIAT Guru)[5] is currently testing whether empowering local communities—by setting up community-school committees and agreeing with teachers on service performance indicators—in combination with different PfP schemes, can lead to better student learning outcomes. Early findings from an impact evaluation of the pilot suggest that PfP schemes can lead to significantly better student learning outcomes and reduced teacher absence (Gaduh et al. 2018).

The favorable findings from the KIAT Guru pilot, together with the ongoing GoI priority to increase education-spending effectiveness, has motivated the GoI to explore the introduction of PfP schemes in urban and semi-urban areas of the country. PfP schemes, however, are only one possible element of a comprehensive teacher evaluation system. For instance, the GoI has introduced the teacher competence test (*Ujian Kompetensi Guru* or UKG) and the teacher performance evaluation (*Penilaian Kinerja Guru* or PKG) in recent years, among many other initiatives, in order to inform employment and salary decisions, as well as teacher professional development and promotion. As the Indonesian teacher evaluation system will likely undergo further reforms in the near future, this paper examines the preferences of key education stakeholders regarding different evaluation methods and indicators currently in use.

Using data from a large-scale opinion survey in Indonesia conducted in 2017, this paper finds that both principals and teachers consider UKG and PKG evaluations as useful methods for improving teacher performance. A majority of respondents stated that these evaluations should occur on an annual basis.

With respect to the feasibility and viability of teacher PfP schemes, our results show that:

1. Overall, principals and teachers support direct linking of the UKG and PKG evaluations to teacher salaries, with most related UKG and PKG indicators registering approval of more than 70 percent in this regard.
2. Teachers strongly favor teacher PfP schemes (97 percent approval rate) over schemes that link salaries to seniority (34 percent approval rate).
3. Overall, this paper finds that teacher PfP schemes are well supported, with the highest level of support coming from teachers who work in urban areas. Multivariate regression analysis shows that teachers in urban areas are 10–13 percentage points more likely to support various PfP schemes compared to teachers in semi-urban areas.
4. Teachers are open to the idea of linking additional indicators outside of the UKG and PKG—such as student learning outcomes—to their professional career path, and therefore to their salaries.
5. There are important differences between principals-teachers and parents-pupils opinions on suitable PfP indicators. For instance, principals and teachers prefer indicators that focus on teacher input, such as lesson plans and preparation for classes, while parents tend to favor indicators that emphasize teacher-parent and teacher-student interactions.
6. The notion of school supervisors (pengawas), principals, teachers, parents, and pupils as performance evaluators is generally supported. However, principals and teachers show significantly greater preference for evaluation roles to be undertaken by supervisors, principals and teachers rather than parents and pupils. Likewise, parents are willing to evaluate teachers on a regular basis using indicators with which they feel most familiar—such as teacher-student interactions, teacher discipline and student learning progress.

The main results clearly show a generalized positive opinion towards PfP schemes. Results, however, should be interpreted with caution due to potential biases inherent to opinion data.

The remainder of this paper is structured as follows. Section 2 describes the instruments of teacher performance evaluation in Indonesia that are relevant for this paper. Section 3 discusses the data and methodology used in this paper. Results are shown in Sections 4 and 5, while Section 6 draws the paper to conclusion.

5 The KIAT Guru pilot has been running since 2016 in the remote rural areas of five districts outside of Java. Please see World Bank (2017) for more details regarding KIAT Guru.

02 Teacher Performance Evaluation in Indonesia

As introduced above, two of the major teacher evaluation tools in the Indonesian education system consist of the teacher competence test (UKG) and the teacher performance evaluation (PKG). This paper examines the opinions of key education stakeholders in Indonesia concerning the use of these two instruments as teacher performance indicators. In addition, this paper examines the views of teachers concerning student learning outcomes and teacher absence.

The UKG is a mandatory test directly measuring the competencies and abilities of teachers. The test focuses on subject knowledge and pedagogical content knowledge. The UKG was first implemented in 2012 as part of the teacher certification process, and was followed with nation-wide implementation in 2015. The UKG is a prerequisite for teacher certification that entitles teachers to a professional allowance. However, once a teacher has achieved certification their UKG score is no longer a determining factor in the level of their salary. Consequent to low test scores in the 2015 UKG, the GoI developed a national teacher professional development program in 2016 aimed at raising the competence of those who failed the test (Ministry of Education and Culture 2016; World Bank 2015).

The PKG measures teacher performance by assessing their personal, social, pedagogical and professional characteristics (Chang et al. 2014; World Bank 2010). The evaluation, which rates teacher performance using a scale ranging from A to D, has traditionally been conducted by school principals on an annual basis, covering 14 competencies using 78 indicators.[6]

While student learning outcomes have not yet been implemented as teacher performance indicators in Indonesia, a standardized student assessment with national coverage already regularly takes place. The National Exam (*Ujian Nasional* or UN) tests students of different grades on subjects such as language, math and science to provide measures of school performance and could, in principle, be adopted and adapted as a measure of teacher performance (UNESCO 2017).

6 Over the years teacher evaluation scores have, however, always remained very high (A or at worst B), while student learning outcomes have stagnated over the past 15-20 years. In order to improve the objectivity of teacher evaluations a unit within the Ministry of Education and Culture, Indonesia (MoEC), has attempted to include evaluators other than principals in the evaluation process—such as teachers, parents, community members and representatives from the private sector. MoEC implemented this proposition in 2,000 secondary schools. The initiative however has not been scaled up due to the complexity of the instrument that covers many indicators, some of which are vague and subject to interpretation.


ROSTER
BELAJAR KELOMPOK
GLEDYSIA LODONG (LEDY)
FLAFIANUS TANGIS (RAFI)
SRILIA JEMIA (SRI)
FORTUNATUS BARUNG (MBOING)
KOLINTUS POGOL (KOLIN)
ROSTER HARIAN

03 Data and Methodology

This paper uses data from an opinion survey that was implemented in 100 Indonesian schools. The survey was implemented during April 2017 by the World Bank in collaboration with the Ministry of Education and Culture, Indonesia (MoEC). The survey took place in 10 districts within five provinces across Indonesia (see Table A1), with participating districts selected in a two-stage process. In the first stage, five districts were purposively selected to represent heterogeneity in terms of geography—comprising the categories of very remote, remote, developing, and developed areas. In the second stage, for each of the five districts initially selected, one neighboring district was also selected. Within each district, 10 schools—three primary schools (SD), three junior secondary schools (SMP), three senior secondary schools (SMA), and one vocational school (SMK)—were selected to represent heterogeneity in terms of student learning outcomes. This heterogeneity is represented by lower performing, average performing, and high performing student learning outcomes within each school category, as measured by the National Exam (UN).[7,8]

The survey was administered to 1,605 individuals comprised of principals, teachers, parents, and pupils. In each of the 100 schools, one principal, five teachers, five parents, and five pupils were interviewed by the survey team.[9] Among teachers, four types were interviewed: certified civil servants (n=193), uncertified civil servants (n=100), certified non-civil servants (n=33), and uncertified non-civil servants (n=177). A share of 60 percent of the teachers sampled work in semi-urban schools, while the remaining 40 percent teach in urban areas. Teachers, parents, and students were selected at random. Among students, pupils of all grades between the 4th and 12th grade were sampled. The survey was administered as face-to-face interviews. Most survey questions consist of Likert scale items that allow for five response options: where one stands for 'strongly disagree', two for 'disagree', three for 'undecided', four for 'agree', and five for 'strongly agree'.

Other survey items asked respondents to choose from a list of categories. For instance, respondents chose who—from among school supervisors, principals, teachers, parents, and pupils—they considered to be the most suitable performance evaluators to measure various teacher competencies. In addition, respondents were asked to select their top five performance indicators in a PfP setting out of a list of 17 teacher competencies. Of these 17 competencies, 14 were based on the core competencies listed in the PKG which refer to teacher characteristics and abilities that concern teacher interaction with students,

7 For vocational schools, schools with average learning outcomes, as measured by the UN score, were selected.

8 From the 100 schools in the sample, 83 were public and 17 were private schools. These numbers are similar to the national shares, which show that 80 percent of schools in Indonesia are public and 20 percent are private.

9 There were infrequent minor deviations from this rule.

parents, and their classroom. In addition, two teacher competencies referring to teacher capacity to improve student learning outcomes, and one related to teacher ability to motivate parents, were added to the list.[10] In the following, this list of 17 indicators is referred as the 'extended PKG list'.

This paper analyzes the opinion data produced using three complementary approaches. It presents the total distribution of answers to various PfP-related statements via descriptive tables and figures, as well as the disaggregated distribution by urban status.[11]

Second, it explicitly tests—using the Mann Whitney Wilcoxon test (MWW)—whether the agreeableness of respondents towards various statements is statistically different by urban status.[12] Third, the analysis uses a Linear Probability Model (LPM) to conduct regression analysis that sheds light on the demographic and institutional correlates of favorability towards PfP schemes.

10 See the full list of teacher competencies in Table A2.

11 This includes the analysis of different subgroups depending upon urban status, civil servant status, public status, and gender as criteria. The disaggregation by urban status is the most informative one, as suggested by the number of responses that are statistically different across the corresponding categories. In a few cases where the the location of schools was not recoverable, the sum of the urban and semi-urban subsamples do not add up exactly to the total sample size, as notable in tables below. The remaining subgroup results are available upon request.

12 The MWW is a well-established non-parametric test that properly handles the ordinal nature of the data (Mann and Whitney 1947). With a significance level of (10) 5 percent, p-values below (0.1) 0.05 suggest that subgroup agreeableness are statistically different from each other.

04 Results: Attitudes Towards Evaluations and Performance Indicators

In this section the paper examines the views of various education stakeholders concerning teacher evaluations and specific teacher performance indicators. Opinions on the UKG, student learning outcomes, and teacher absenteeism are discussed. While responses of principals, teachers, parents, and pupils are investigated, the paper focuses strongly on teacher respondents. Throughout, this section focuses upon the full survey sample, and the urban and semi-urban subsamples.[13]

As described in section 2, the UKG is one of the main teacher evaluation schemes and performance indicators that MoEC has introduced over the last decades (Chang et al. 2014; World Bank 2010). Given the experience of principals and teachers with the UKG scheme, this section reviews their opinions of this performance indicator in a general context. Furthermore, several alternative indicators that can be used to evaluate teachers on a regular basis, such as student learning outcomes and teacher absence, are also examined. Regarding these latter indicators, the opinions of various stakeholders (principals, teachers, parents and students) are discussed.

UKG Test

Overall, teacher responses reveal strong support for the UKG as a suitable performance indicator (see Panel A–C of Table 1). First, more than 80 percent of teachers express support for the UKG as a performance assessment tool. In the same vein, more than 72 percent of teachers believe that the UKG can assess their teaching competence. Correspondingly, only 10 percent of teacher respondents believe that the UKG is not useful for career development, further revealing the extent of teacher support for this competency test.

Second, teacher responses hint at the suitability of the UKG as a performance indicator in different ways (see Panel D and E of Table 1). For instance, its regular use is supported by the majority of teachers—a share of 71 percent favors the idea of undertaking the UKG on an annual basis. Responses that concern the difficulty of the UKG test also show its viability as a performance indicator, as the perceived difficulty is not concentrated at the tail end of the scale. When asked about the difficulty of the UKG on a rating scale using five categories ranging from 'very hard' to 'very easy', almost 40 percent of teacher respondents reported a difficulty of middle-range, while 50 percent believe the UKG is 'hard'. Moreover, the MWW test suggests that teachers in urban areas demonstrate systematically higher levels of support for the UKG as a performance indicator than teachers in semi-urban schools.

13 Geographical areas of Indonesia are administratively categorized into cities and districts. Under the district category, further division is based on the Developing Villages Index (***Index Desa Membangun***) which identifies developed villages, developing villages, disadvantaged villages, and very disadvantaged villages. The urban sample in this group includes cities and developed villages, while the semi- urban sample includes developing villages.

Table 1. Teachers: UKG as Performance Indicators

Panel A. Statement: 'UKG should be linked to teacher performance assessment'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	9.8	2.4	71.3	16.5	87.8
Semi-urban	0.0	16.3	7.2	64.6	11.8	76.4
Total	0.0	13.8	5.4	67.1	13.8	80.9
p-value	.01					

Panel B. Statement: 'UKG is able to assess your competence'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	15.2	3.0	62.2	19.5	81.7
Semi-urban	1.5	24.0	8.0	58.6	8.0	66.5
Total	0.9	20.5	6.1	59.9	12.6	72.5
p-value	0					

Panel C. Statement: 'UKG is not useful for career development'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	16.5	70.1	4.9	6.7	1.8	8.5
Semi-urban	11.0	71.5	6.5	10.6	0.4	11.0
Total	13.3	70.9	5.8	9.1	0.9	10.0
p-value	.08					

Panel D. Statement: 'How difficult is UKG?'

	Very Hard	Hard	Neutral	Easy	Very Easy
Urban	6.1	49.7	36.7	7.5	0.0
Semi-urban	5.9	49.6	41.9	2.5	0.0
Total	6.0	49.9	39.7	4.4	0.0
p-value	.74				

Panel E. Statement: 'How often should UKG be implemented? Every...'

	1 year	2 years	3 years	4 years	5 years
Urban	80.9	13.8	2.6	0.0	2.6
Semi-urban	64.2	21.8	10.0	0.4	3.5
Total	71.0	18.5	7.0	0.3	3.1
p-value	0				

Note: Panel A–C have a teacher sample of 429 observations, of which 164 are urban and 263 semi-urban. Panel D has a teacher sample of 385 observations, of which 147 are urban and 236 semi-urban. Values are in percentages. Panel E has a teacher sample of 383 observations, of which 152 are urban and 229 semi-urban. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

Table 2 shows that the majority of principals who were respondents also support the UKG as a performance indicator. A share of 78 percent of principals believe the UKG should be linked to the teacher performance assessment. Moreover, 69 percent of principals think it is also well suited to assess teacher competence. In line with teachers in urban areas, principals in urban areas are systematically more favorable to these two statements than principals in semi-urban areas. Moreover, a share of 71 percent of principals agree with conducting the competence test on an annual basis. Interestingly, 72 percent of principals reported to agree or strongly agree with the notion that the UKG forces teachers to improve their competencies.

Table 2. Principals: UKG as Performance Indicators

Panel A. Statement: 'UKG should be linked to teacher performance assessment'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	7.5	5.0	70.0	17.5	87.5
Semi-urban	0.0	23.3	5.0	66.7	5.0	71.7
Total	0.0	17.0	5.0	68.0	10.0	78.0
p-value	.01					

Panel B. Statement: 'UKG forces teachers to improve competence'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	5.0	30.0	0.0	45.0	20.0	65.0
Semi-urban	0.0	21.7	1.7	56.7	20.0	76.7
Total	2.0	25.0	1.0	52.0	20.0	72.0
p-value	.31					

Panel C. Statement: 'UKG is well suited to assess teacher competence'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	17.5	7.5	35.0	40.0	75.0
Semi-urban	0.0	25.0	10.0	46.7	18.3	65.0
Total	0.0	22.0	9.0	42.0	27.0	69.0
p-value	.05					

Panel D. Statement: 'How often should UKG be carried out. Every...'

	1 year	2 years	3 years	4 years	5 years
Urban	76.3	13.2	7.9	2.6	0.0
Semi-urban	67.3	14.5	9.1	1.8	7.3
Total	71.0	14.0	8.6	2.2	4.3
p-value	.28	.28			

Note: KIAT Guru Urban Opinion Survey 2017. Panel A–C have a principal sample of 100 observations, of which 40 are urban and 60 semi-urban. Panel D has a principal sample of 93 observations, of which 38 are urban and 55 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

Student Learning Outcomes

In most countries, including Indonesia, teacher evaluations are usually linked to education inputs such as presence, pedagogical skills, teaching skills, and so forth. However, in some countries teacher evaluations are more directly linked to education outputs, such as student learning outcomes. Intuitively, output-oriented teacher performance indicators should be measures that teachers can influence and have a direct impact upon. Therefore, it is critical to identify whether teachers believe that they can directly influence student learning. As shown in Table 3 and Table 4, the majority of teachers surveyed are confident in being able to overcome student learning barriers unrelated to teachers, such as limitations in the financial background or home environment of a student, as well as poor preparation from previous grades, among other potential barriers. In general, these responses imply that student learning outcomes are perceived to depend upon teachers' abilities, and hence indirectly support this indicator as a performance measure.[14]

Table 3. Teachers: Teachers Influence and Student Profiles (Table 1)

Panel A. Statement: 'Little I can do to help students learn if parents do not seek feedback from teachers'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	3.5	50.0	7.0	34.5	5.0	39.5
Semi-urban	5.7	54.0	9.7	27.3	3.3	30.7
Total	5.0	52.5	8.5	30.0	4.0	34.0
p-value	.06					

Panel B. Statement: 'Little I can do to help students learn if students come unprepared from previous grades'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	5.0	56.0	7.5	27.0	4.5	31.5
Semi-urban	8.7	59.7	8.7	20.3	2.7	23.0
Total	7.4	58.3	8.2	22.9	3.4	26.2
p-value	.03					

Panel C. Statement: 'Little I can do to help students learn if parents have too many problems to be concerned with the child's education'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	11.0	54.0	7.5	19.5	8.0	27.5
Semi-urban	9.7	57.0	8.3	22.7	2.3	25.0
Total	10.3	55.9	8.0	21.3	4.6	25.8
p-value	.62					

14 In line with this, 65 percent of parents believe that their child's learning outcomes are the product of their teacher's ability to teach (see Figure A3 in the Appendix).

Panel D. Statement: 'Little I can do to help students learn if students come unprepared to do school works'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	7.5	54.0	4.5	28.5	5.5	34.0
Semi-urban	10.7	65.3	4.7	17.0	2.3	19.3
Total	9.5	60.8	4.6	21.5	3.6	25.0
p-value	0					

Panel E. Statement: 'Little I can do to help students learn if parents do not have the necessary education to help the child'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	6.5	63.5	5.5	22.5	2.0	24.5
Semi-urban	9.7	74.7	4.0	10.3	1.3	11.7
Total	8.5	70.2	4.6	15.1	1.6	16.7
p-value	0					

Source: KIAT GURU Urban Opinion Survey 2017. Teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

The subgroup analysis in Table 3 suggests that a higher share of teachers in urban schools believe they are capable of helping disadvantaged students than do teachers in semi-urban schools. However, when it comes to the specific belief that teachers are able to overcome the influence of the home environment on student performance, semi-urban teachers seem to systematically agree they are more able to do so compared to urban teachers, as shown in Panel E of Table 4 Table 4.

Table 4. Teachers: Teachers Influence and Student Profiles (Table II)

Panel A. Statement: 'I am confident I can motivate students to learn regardless of their financial status'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	5.0	1.0	43.5	50.5	94.0
Semi-urban	1.3	2.0	1.3	46.7	48.7	95.3
Total	0.8	3.2	1.2	45.5	49.3	94.8
p-value	.81					

Panel B. Statement: 'I am confident I can compensate for the poor preparation of some students from previous grades'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.5	1.0	3.0	69.5	26.0	95.5
Semi-urban	0.7	0.7	6.0	69.3	23.3	92.7
Total	0.6	0.8	4.8	69.4	24.5	93.8
p-value	.28					

Panel C. Statement: 'I am confident I am able to help even the lowest performing students'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	2.0	7.0	2.5	58.0	30.5	88.5
Semi-urban	0.7	1.3	2.7	62.0	33.3	95.3
Total	1.2	3.6	2.6	60.4	32.2	92.6
p-value	.1					

Panel D. Statement: 'I am held responsible for my students' learning outcomes even though their learning process is influenced by many factors'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.0	16.5	6.5	55.0	21.0	76.0
Semi-urban	0.3	13.7	4.7	57.3	24.0	81.3
Total	0.6	14.7	5.4	56.5	22.9	79.3
p-value	.17					

Panel E. Statement: 'I am confident I can overcome the influence of the home environment on student performance'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.5	8.0	16.0	63.5	11.0	74.5
Semi-urban	1.0	9.7	23.0	58.7	7.7	66.3
Total	1.2	8.9	20.3	60.6	8.9	69.6
p-value	.05					

Note: KIAT Guru Urban Opinion Survey 2017. Teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

While the previous tables have shown that teachers feel capable of shaping learning outcomes of disadvantaged students, they do not inform us as to whether teachers believe disadvantaged students deserve more of their attention. Teachers in medium and low-income settings might face classrooms exhibiting significant discrepancies between students' abilities and needs (World Bank 2018b). In such contexts it might be difficult for teachers to pay equal attention to all students. In line with this scenario, two-thirds of teachers interviewed believe it is difficult for them to pay equal attention to all students within a large classroom. Moreover, the share of teachers in semi-urban schools who share this perspective is 10 percentage points higher than teachers in urban schools (see Table 5).

The majority of teachers responded that advantaged students deserve more of their attention than disadvantaged students. According to the large majority of teachers, students whose parents are involved and willing to invest in their child's education deserve more teacher attention than other students. The same applies for students that are more motivated to learn, attend school regularly, come to school with the materials necessary to complete school work, have the necessary foundation from previous classes, and perform well in class. However, while teacher opinions predominantly indicate that more attention should be given to 'good' students, teachers also expressed opinions that students who lag behind in classwork or homework also deserve more of their attention. At the same time, teachers believe they are capable of shaping the learning outcomes of disadvantaged students, as shown in Table 3 and Table 4.

In summary, most teachers favor the idea of providing additional attention to better-performing students, a finding that has been observed in other low and middle-income contexts (World Bank 2018b; Sabarwal and Abu-Jawdeh 2017; Abadzi and Llambiri 2011). There may

be various explanations for this behavior; for example, high-ability students are easier to teach and might provide immediate teaching satisfaction. Likewise, teachers might believe that their provision of additional learning support is a fair reward for the good performance of motivated students.

It is difficult to predict what consequences would result from a scenario of increased teacher effort induced by teacher evaluation. The direction of any effect on the ability gap would depend upon how teachers allocate additional attention across students with different profiles, and upon the nature of marginal returns to teacher attention.

Table 5. Teachers: Heterogeneous Attention

Panel A. Statement: 'It is difficult for me to pay equal attention to all my students in a large class'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	8.0	31.5	0.5	43.5	16.5	60.0
Semi-urban	2.0	26.7	0.7	53.0	17.7	70.7
Total	4.4	28.8	0.6	49.1	17.1	66.2
p-value	.02					

Panel B. Statement: 'Students deserve more of my attention if they are performing well in class'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	4.0	3.0	66.0	27.0	93.0
Semi-urban	0.3	5.3	1.0	62.7	30.7	93.3
Total	0.2	5.0	1.8	63.6	29.4	93.0
p-value	.43					

Panel C. Statement: 'Students deserve more of my attention if they are lagging behind in classwork/ homework'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	2.0	13.0	3.0	50.0	32.0	82.0
Semi-urban	0.7	6.3	0.3	49.7	43.0	92.7
Total	1.2	8.9	1.4	49.7	38.8	88.5
p-value	0					

Panel D. Statement: 'Students deserve more of my attention if they have the necessary foundation from previous classes'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.0	12.0	4.5	59.5	23.0	82.5
Semi-urban	0.3	10.3	1.7	68.0	19.7	87.7
Total	0.6	11.1	2.8	64.6	20.9	85.5
p-value	.82					

Panel E. Statement: 'Students deserve more of my attention if they are motivated to learn'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	3.5	15.0	2.5	56.0	23.0	79.0
Semi-urban	1.0	13.7	2.3	50.0	33.0	83.0
Total	2.0	14.3	2.4	52.5	28.8	81.3
p-value	.02					

Panel F. Statement: 'Students deserve more of my attention if they come to school with the material necessary to do school work'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.0	14.5	1.5	60.5	22.5	83.0
Semi-urban	0.7	16.7	2.7	60.3	19.7	80.0
Total	0.8	15.7	2.2	60.6	20.7	81.3
p-value	.34					

Panel G. Statement: 'Students deserve more of my attention if they are attending school regularly'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	2.5	21.0	2.0	56.5	18.0	74.5
Semi-urban	0.7	13.7	2.0	59.7	24.0	83.7
Total	1.4	16.7	2.0	58.4	21.5	79.9
p-value	.01					

Panel H. Statement: 'Students deserve more of my attention if parents are involved in the education of their child'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.0	14.5	4.5	53.5	26.5	80.0
Semi-urban	0.7	18.7	3.3	58.0	19.3	77.3
Total	0.8	17.3	3.8	56.1	22.1	78.1
p-value	.09					

Panel I. Statement: 'Students deserve more of my attention if parents are willing to invest the necessary financial resources in the education'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	4.0	26.0	8.5	42.5	19.0	61.5
Semi-urban	4.7	41.0	11.7	34.3	8.3	42.7
Total	4.4	34.8	10.5	37.8	12.5	50.3
p-value	0					

Note: Teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements of Panels B-I are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

Two survey items collected direct opinions from principals and teachers respectively on the use of student learning outcomes as a teacher performance indicator, which received relatively strong support. As depicted in Table 6, around 70 percent of principals and teachers agree that student test scores should be the main factor in assessing teacher performance. Interestingly, teachers of urban schools have a systematically higher favorability towards this statement, as suggested by the MWW test.

This suggests that student learning outcomes are able to proxy for the relevant set of indicators selected as best-teacher-performance indicators.[16] Overall, responses indicate that student test scores have relatively strong support as a teacher performance indicator.

Teacher Absenteeism

Teacher presence in school and class is another performance indicator that can be linked to teacher evaluations. It is well-documented that teachers

Table 6. Principals and Teachers: Student Test Scores as Performance Indicator

Panel A (Principals). Statement: 'Main indicator for teacher performance should be students' test scores'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	25.0	2.5	42.5	30.0	72.5
Semi-urban	3.3	23.3	3.3	51.7	18.3	70.0
Total	2.0	24.0	3.0	48.0	23.0	71.0
p-value	.32					

Panel B (Teachers). Statement: 'Main indicator for teacher performance should be students' test scores'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	13.5	3.0	68.5	15.0	83.5
Semi-urban	0.7	31.0	7.3	51.7	9.3	61.0
Total	0.4	24.3	5.6	58.3	11.5	69.8
p-value	0					

Note: Panel A: Principal sample of 100 observations, of which 40 are urban and 60 semi-urban. Panel B: Teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis) agree'. Reported p-values correspond to the MWW test.

Notably, the support for student learning outcomes as a teacher performance indicator was somewhat weaker in comparison to the top five indicators teachers chosen from the 'extended PKG list' to assess teacher performance; the top five indicators of key education stakeholders are examined in detail in Section 5.[15] From the extended PKG list, about 30 percent of teachers selected improvements in subject-specific learning outcomes, while principals showed slightly stronger support than teachers for student learning outcomes as a teacher performance indicator. Importantly, four of the five teacher competencies chosen by teachers as most important to assessing teacher performance were also chosen by them as the most important factors for student learning outcomes.

in Indonesia are often absent (ACDP 2014) despite teacher-specific presence indicators being routinely collected by district education offices. The problem with teacher presence indicators lies with the absence of accurate data concerning teacher presence, with reported presence rates almost always indicating 100 percent presence.[17]

Furthermore, teachers often seem to find teacher absence quite acceptable. A substantial share—although

15 The extended PKG list of teacher competencies for teacher performance evaluation is shown in Table A2 of the Appendix.

16 The four indicators chosen in both questions are whether teachers: have a strong work ethic, sense of responsibility, and sense of professional pride; can translate the curriculum into lesson plans; have mastered educative teaching and learning theory and principles; and have mastered their subject.

17 In the KIAT Guru pilot impact evaluation, tying teacher remote area allowances with teacher presence significantly improves time spent in teaching, parental involvement, and student learning outcomes (Gaduh et. al. 2018). Teacher presence is documented daily using an Android-based application, and verified monthly by community and parent representatives. The tamper-proof and verifiable evidence that is produced provides an objective measure that makes it difficult for teachers to shirk.

not the majority—of teachers justify teacher absence if certain conditions are met, as shown in Table 7. For instance, 35 percent of teachers think it is acceptable to be absent from teaching if they leave students with work to do, or if teachers have completed their assigned curriculum. Similarly, more than 28 percent of teachers justify absenteeism if the tasks they carry out during their absence are useful for the community.[18] These numbers indicate that a significant share of teachers do not perceive absenteeism as consciously shirking, but as a justifiable and acceptable practice under specific conditions.

92 percent of parents agree with the statement that teachers go to school and teach regularly. In other words, the vast majority of parents believes that there is a rather low rate of teacher absenteeism in general. On the other hand, almost 31 percent of the students responded that teachers often do not start and end the class on time. Similarly, more than a quarter of students responded that their teachers are often not present for the entire duration of a lesson. Hence, student responses hint at a substantial rate of teacher absenteeism (see Figure A1 and Figure A2 in the Appendix). Taken together, these

Table 7. Teachers: Acceptability of Teacher Absenteeism

Panel A. Statement: 'I think it is acceptable for me to be absent if I leave students with work to do in my absence'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	15.0	48.0	6.0	27.0	4.0	31.0
Semi-urban	9.0	44.3	8.7	36.3	1.7	38.0
Total	11.3	46.1	7.6	32.4	2.6	35.0
p-value	.03					

Panel B. Statement: 'I think it is acceptable for me to be absent if I complete my assigned curriculum'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	14.0	48.0	5.0	28.0	5.0	33.0
Semi-urban	9.7	44.7	9.3	32.3	4.0	36.3
Total	11.3	46.1	7.6	30.4	4.6	35.0
p-value	.13					

Panel C. Statement: 'I think it is acceptable for me to be absent if I am doing something useful for the community'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	10.5	57.0	8.5	23.0	1.0	24.0
Semi-urban	8.7	46.3	13.3	28.7	3.0	31.7
Total	9.5	50.7	11.3	26.2	2.2	28.4
p-value	.01					

Note: Teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

In addition, parents and students were asked about teacher behavior related to the prevalence of teacher absenteeism. Interestingly, responses by parents and students give a somewhat different picture than that provided by teachers. On the one hand, more than

responses suggest that teachers are often present at school but often absent from class; a result in line with the latest figures from teacher absenteeism surveys (McKenzie et al. 2014; UNCEN et al. 2012). An alternative interpretation of these results would be that student responses are simply more informed than parent responses with respect to teacher absence.

18 Such numbers place Indonesia in the middle range of absenteeism acceptability among the eight countries analyzed by Sabarwal and Abu-Jawdeh (2017).

05 Results: Attitudes Towards High-Stakes Evaluations

Education stakeholders' choice of suitable indicators, and people suitable to be evaluators, can differ significantly depending on whether an evaluation affects teacher salaries or not. Particularly in high-stakes evaluations, such as those affecting teacher promotion and career (e.g. becoming a civil servant or becoming certified) or teacher salaries (e.g. PfP schemes in KIAT Guru), schemes need to be carefully designed. The success of PfP schemes (and their incentive mechanisms) rely heavily on the compliance of service providers, which is dependent upon providers' opinions of such schemes. This section examines the views of education stakeholders concerning high-stakes evaluations, with an emphasis on PfP schemes.

Key Teacher Performance Evaluation (PKG) indicators

Education stakeholders were asked to list and rank up to five indicators they feel are most important for achieving better teacher performance and that should be linked to teacher PfP schemes. To limit the number of indicator choices, respondents were asked to select from the 17 items that comprise the extended PKK competency list. The ranking reported below was determined by the frequency of indicators chosen by each respondent type. As shown in the next table, principals and teachers exhibited similar attitudes in their assessment. Moreover, their preferences often differed from the preferences of parents.

On the one hand, principals and teachers prioritized indicators that focused on the teacher alone. Both suggested the following indicators as one of their top five, whether a teacher has a strong work ethic, sense of responsibility, and sense of professional pride; can develop curriculum into lesson plans; and continuously improves their teaching competence, knowledge and skills. In addition, teachers believe that mastering their subject is a good indicator. Principals, however, more often referred to the capacity of teachers to improve student learning outcomes, and their capacity to conduct teaching and learning activities, as the best performance indicators.

Parents on the other hand, more often chose indicators that reflect teacher-parent/student interaction and communication skills. The two most frequently chosen indicators by parents refer to the ability of teachers to assess the characteristics of a student, and whether teachers are able to communicate with other teachers, parents, students, education personnel, and the community. Parents also reported that teachers should behave in line with moral, social, cultural, and religious norms as an important indicator. The 4th and 5th competencies most commonly chosen by parents refer to a teacher's capability to teach, such as conducting teaching and learning activities, and mastering educative teaching and learning theory and principles.

Table 8. Ranking of Teacher Competencies That Shall Influence Teacher Salaries

	Teachers	Principals	Parents	Combined
Teachers				
Have strong work ethic, sense of responsibility, and sense of professional pride	1	1	-	1
Can develop curriculum into lesson plans	2	2	-	2
Continuously improve their teaching competence, knowledge, and skills	3	4	-	5
Master educative teaching and learning theory and principles	4	3	5	4
Master their subject matter	5	-	-	-
Improve learning outcomes	-	5	-	-
Can assess students' characteristics	-	-	1	3
Able to communicate with teachers, parents, education personnel, students, and the community	-	-	2	-
Behave in line with moral, social, cultural, and religious norms	-	-	3	-
Conduct teaching and learning activities	-	5	4	-

Note: Sample of 488 teacher observations, 64 principal observations and 488 parent observations. Indicators are in descending order after teacher responses. Only the top five indicators for each type of respondent are included. See Table A2 in the Appendix for the full list of indicators.

Who Should Evaluate Teachers?

A performance evaluation is a complex process that requires a certain comfort level, mutual trust, and the respect and acceptance of both the evaluator and the evaluated. Consequently, shared stakeholder outlooks on the suitability of potential evaluators are fundamental to discussion and design of future policy measures concerning evaluations. Currently, teacher performance in Indonesia is evaluated by principals (Chang et al. 2014; World Bank 2010). This section reviews the opinions of education stakeholders (principals, teachers, parents and students) concerning issues related to suitable evaluators for teacher performance assessments.

Principals and teachers were asked who—out of a list of five different education stakeholders—they thought could provide an accurate assessment of the five selected PfP performance indicators. Choices for evaluators consisted of school inspectors, principals, teachers, parents and pupils. Interestingly, principals and teachers hold a very similar attitude towards the suitability of evaluators for high-stakes performance indicators (see Table 9), both showing the greatest support for principals. Over 80 percent of principal and teacher respondents believe other teachers and school inspectors are also well suited to be evaluators of key performance indicators. Notably, principals and teachers gave pupils a 10 percentage point lead over parents, with shares above 60 percent.

Students and parents were asked how comfortable they felt as an evaluator of teacher performance. Results in Figure 1 show that parents are generally comfortable with the idea of evaluating teacher performance when the evaluation influences pay and promotion. The majority of parents reported feeling comfortable evaluating each of the teacher competencies on the extended PKG list of 17 indicators; never more than 28 percent of parents indicated feeling uncomfortable evaluating any particular competency. The indicators parents feel most comfortable in evaluating are: able to communicate with teachers, parents, education personnel, students,

Table 9. Teachers and Principals: Attitudes Towards Stakeholders as Evaluators

Panel A. Sum of shares of agree and strongly agree with following stakeholder as evaluator (%)

	Pengawas	Principals	Other Teachers	Parents	Pupils
Teachers	81.2	95.9	83.0	52.9	62.9
Principals	85.4	100	85.4	51.2	60.0

Note: Principals were also asked whether parental assessments should be part of teacher performance evaluation. 78% of the respondents agree or strongly agree with that statement. Teacher sample of 503 observations. Principal sample of 100 observations. To calculate the values shown, the shares of agree and strongly agree for the evaluator questions involving the top five teacher performance indicators chosen by each respondent are added up. In a second step, the average over these five values is calculated.

and the community (71 percent); and the capacity of teachers to develop the potential of their students (71 percent). In contrast, parents were relatively less willing to assess whether: a teacher is a role model (58 percent), whether teachers master their subject (57 percent), and whether teachers continually improve their competence, knowledge and skills (56 percent).

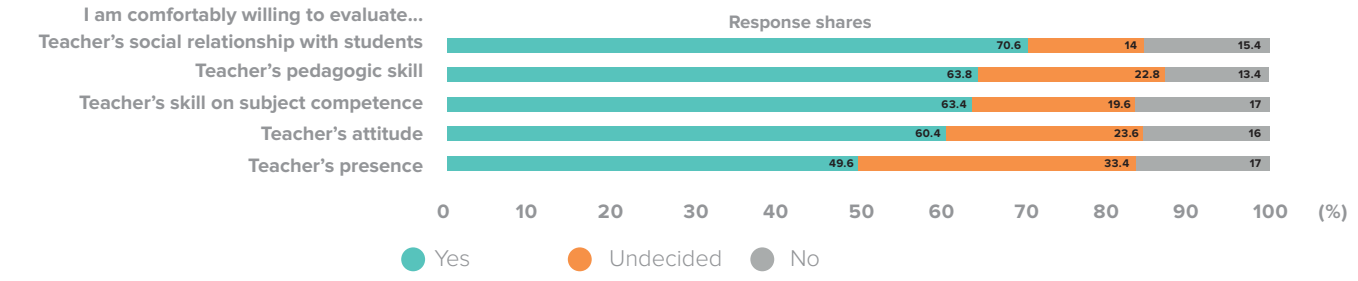
Pupils were asked, using a shorter list of indicators than those provided to parents, how comfortable they felt as an evaluator of teacher performance (see Figure 2 below). The majority of pupils felt comfortable and willing to evaluate their teachers regarding most of the indicators provided, although this question was not asked in a PfP setting.[19] Pupils felt particularly capable of evaluating the social relationship between teachers and students, as well as a teacher's pedagogic skills.

Intriguingly, the category that received least approval—less than half of student respondents—concerns the evaluation of teacher presence. A potential explanation consistent with this large share of indecisiveness involves well known courtesy biases in reporting, whereby students feel uneasy about reporting the absence of their teacher. Evaluating teacher absence may prove more compromising for students than evaluating other performance indicators. As reporting teacher absence is hard evidence indicating a serious lack of teacher effort, with potentially severe consequences for the teacher, student evaluators who sympathize with their teachers might find themselves in a compromising situation they would prefer to avoid. Furthermore, pupils may be afraid of retaliation by teachers in the case of unfavorable evaluations.

Figure 1. Parents: I am comfortable providing an assessment of the following teacher skills/characteristics as performance indicators that would influence their payment

Note: Parent sample with varying number of observations (74–302) depending on competency.

Figure 2. Students: Comfortably willing to evaluate teacher competencies



Note: Student sample of 500 observations.

19 It is unclear whether children would have understood the concept of a PfP setting. For students, this question did not explicitly refer to either payment or promotion consequences of the evaluation.

Pay Criteria

Teachers were asked for their opinion on whether their salary should be linked to their performance or their seniority. Results show that teachers overwhelmingly prefer their payment to be linked to teacher performance over seniority.

As depicted in Table 10, *almost all respondents* agree or strongly agree with the idea of having teacher promotions—which typically affect their payments—dependent upon teacher performance. Likewise, most teachers agree that their salary be based on teacher performance assessments. In contrast, the majority of teacher respondents reject the idea of linking teacher promotions or salaries to seniority. While overall support amongst teachers for linking teacher promotion and salary to seniority is low, teachers in urban schools show a systematically higher level of favorability towards seniority.

Results indicate that teachers consider strong teacher support for the UKG, PKG and student learning outcomes as appropriate performance based evaluation indicators to link to teacher salaries (Table 11). Over 83 percent of

Table 10. Teachers: Seniority and Teacher Performance as Pay Criteria

Panel A. Statement: 'Teacher promotion should be based on teacher performance'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.5	0.5	2.0	66.0	31.0	97.0
Semi-urban	0.3	2.0	0.3	65.3	32.0	97.3
Total	0.4	1.4	1.0	65.6	31.6	97.2
p-value	.79					

Panel B. Statement: 'PKG should affect the teacher's salary'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.5	18.5	4.5	65.5	10.0	75.5
Semi-urban	2.0	21.0	8.7	60.0	8.3	68.3
Total	1.8	19.9	7.2	62.0	9.1	71.2
p-value	.12					

Panel C. Statement: 'Teacher promotion should be based on seniority'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	5.0	44.5	7.0	40.5	3.0	43.5
Semi-urban	7.3	56.7	8.0	26.3	1.7	28.0
Total	6.6	51.7	7.6	32.0	2.2	34.2
p-value	0					

Panel D. Statement: 'Teacher salary should be linked to seniority'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	10.0	50.0	4.5	31.0	4.5	35.5
Semi-urban	10.7	59.0	10.3	19.0	1.0	20.0
Total	10.5	55.5	8.0	23.7	2.4	26.0
p-value	.01					

Note: Teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

teachers believe the UKG should be part of the teacher certification process. Moreover, 62 percent believe that it should also be linked to the TPG payment. The PKG and student learning outcomes are also strongly supported by teachers as indicators suitable to influence salary and promotion, respectively. It should be noted that the UKG and PKG receive greater support from teachers (and principals) than does student learning outcomes. Moreover, teachers in urban schools express systematically higher support for the UKG and student learning outcomes as appropriate performance based evaluation indicators linked to teacher salaries than do teachers in semi-urban schools, as shown in Panel A, C and D.

Teachers demonstrate a different opinion on student learning outcomes depending on the type of pay component in question (Table 11). While the majority of teacher respondents believe student learning outcomes should influence teacher promotion (see Panel D), only 17 percent favor the idea of receiving a bonus as a result of good student learning outcomes. When comparing these results with opinion surveys in other countries, a similar rejection of the bonus scheme is observed in Argentina. In contrast, country samples from Afghanistan, India, Myanmar, Pakistan, Senegal, Tajikistan, and Tanzania indicate strong teacher support for payment schemes that reward teachers with bonuses for good student learning outcome results (Sabarwal and Abu-Jawdeh 2017; Muralidharan and Sundararaman 2011a).

Table 11. Teachers: Pay Criteria

Panel A. Statement: 'UKG should be part of the teacher certification process'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.6	7.3	2.4	71.3	18.3	89.6
Semi-urban	0.4	13.7	6.1	68.1	11.8	79.8
Total	0.5	11.2	4.7	69.2	14.5	83.7
p-value	0					

Panel B. Statement: 'PKG should affect the teacher's salary'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	1.5	18.5	4.5	65.5	10.0	75.5
Semi-urban	2.0	21.0	8.7	60.0	8.3	68.3
Total	1.8	19.9	7.2	62.0	9.1	71.2
p-value	.12					

Panel C. Statement: 'UKG should be linked to TPG'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.6	21.3	4.9	56.7	16.5	73.2
Semi-urban	0.8	33.1	11.0	45.2	9.9	55.1
Total	0.7	28.4	8.6	49.7	12.6	62.2
p-value	0					

Panel D. Statement: 'My promotion should partly depend on my students' test scores'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	3.5	20.5	6.5	59.0	10.5	69.5
Semi-urban	1.7	30.0	12.0	49.3	7.0	56.3
Total	2.4	26.2	9.7	53.3	8.3	61.6
p-value	.01					

Panel E. Statement: 'If my students perform well in exams I should receive a bonus'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	21.0	48.5	5.0	19.0	6.5	25.5
Semi-urban	14.0	67.0	7.7	9.0	2.3	11.3
Total	16.7	59.8	6.6	12.9	4.0	16.9
p-value	.22					

Note: Panel A and C have a teacher sample of 429 observations, of which 164 are urban and 263 semi-urban. Panel B, D, and E have a teacher sample of 503 observations, of which 200 are urban and 300 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

In sum, teachers support the idea of PfP schemes. A potential source of popularity for these schemes could be the high levels of perceived fairness and transparency that teachers report concerning existing elements of the teacher performance assessment process: such as the PKG process, the teacher certification process, teacher promotions, and workload divisions. Such high levels of perceived fairness and transparency within the system are likely to foster teacher trust in the reliability of system administrators, and may motivate teachers to accept performance-linked pay. Muralidharan and Sundararaman (2011a), who used a mixed methods approach in India, point to this explanation. A second source of the popularity among teachers of student learning outcomes as a basis for pay schemes, is teacher belief that they are able to influence student scores, as discussed above. The large majority of teachers expressed confidence in their capacity to influence student scores, including the scores of students with disadvantaged profiles. In line with this result, teachers

Table 12. Principals: Specific Pay Criteria

Panel A. Statement: 'UKG should be part of the teacher certification process'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	7.5	2.5	77.5	12.5	90.0
Semi-urban	1.7	16.7	8.3	63.3	10.0	73.3
Total	1.0	13.0	6.0	69.0	11.0	80.0
p-value	.08					

Panel B. Statement: 'Student test scores should be considered in teacher promotion'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	2.5	17.5	7.5	52.5	20.0	72.5
Semi-urban	0.0	11.7	3.3	63.3	21.7	85.0
Total	1.0	14.0	5.0	59.0	21.0	80.0
p-value	.28					

Panel C. Statement: 'UKG should be linked to TPG'

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree	Total Agree and Strongly Agree
Urban	0.0	15.0	7.5	70.0	7.5	77.5
Semi-urban	0.0	31.7	10.0	50.0	8.3	58.3
Total	0.0	25.0	9.0	58.0	8.0	66.0
p-value	.09					

Note: Principal sample of 100 observations, of which 40 are urban and 60 semi-urban. Values are in percentages. 'Total (dis)agree' is calculated as the sum of 'Strongly (dis)agree' and '(dis)agree'. Statements are shown in descending order after values of 'Total agree'. Reported p-values correspond to the MWW test.

agree with being held accountable for student learning outcomes.

Finally, the survey asked principals about the role of the UKG and student learning outcomes in affecting teacher salaries. As shown in Table 12, principals' responses correlate with teachers' opinions. The large majority of principals favor the UKG (as a criterion for teacher certification) and TPG, while student test scores are supported as a valid determinate in teacher promotion.

Regression Analysis

The statements presented in Table 13 and Table 14 are particularly relevant for policy considerations. While they show that PfP schemes are generally well supported by teachers, certain teacher characteristics are associated with higher or lower levels of support. Therefore, this paper conducted a multivariate regression analysis to shed light on the demographic and institutional correlates of teacher agreeableness on survey statements.[20] Table 13 and Table 14 show the results of LPM regressions with the dependent variable taking the value of one if the teacher strongly agrees with the statement and zero otherwise.[21]

The regression framework investigates whether demographic factors (such as being a female teacher, age, having a Bachelor of Education degree or higher, or having passed the teacher certification process) are systematically related to higher support for PfP-related statements. The regressions also control for institutional factors, such as whether a teacher is a civil servant, and whether a teacher works at a public school. Finally, a controlling binary indicator is considered for whether a school is located in a semi-urban area as opposed to a fully urbanized area.[22]

Table 13 shows regressions on seniority and teacher performance as criteria for pay. The first two columns show that none of the listed factors are systematically related to higher support for linking teacher performance to teacher salaries or for relating the PKG to teacher salaries.

For the first association, this result is not surprising. Given that more than 97 percent of teachers agree with the statement that teacher promotion should be based on teacher performance, there is almost no variation to be explained by any potential predictor. On the contrary, regressions involving seniority as a criterion for promotion and salary show statistically significant results. The probability of supporting seniority increases with age, while it decreases if the teacher is a civil servant or if the school is located in a semi-urban area. In addition, female teachers are more likely to support the idea of linking salary to seniority, while more educated teachers are less likely to support this idea. Moreover, the magnitude of these effects is considerable. For instance, being a civil servant reduces the probability of supporting seniority as a criterion for teacher promotion by 21 percentage points, while the probability of a 50 year old teacher supporting seniority as a criterion for teacher promotion is 20 percentage points higher than for a 30 years old teacher.

Table 14 presents the regression results on the teacher agreeableness for various PfP schemes involving the UKG, PKG and students' test scores. Noticeably, teachers of semi-urban schools are less likely to favor PfP systems in four out of five proposed schemes. For instance, teachers of semi-urban schools have a 10 percentage points lower probability of supporting the inclusion of the UKG in the teacher certification process. Interestingly, the magnitude of the coefficient remains similar across the different schemes. This suggests that when it comes to teacher acceptance, the implementation of PfP schemes might be less challenging in fully urbanized areas as compared to semi-urbanized ones. Finally, certified teachers are less likely to support the idea of linking the UKG to TPG.

20 This exercise was not undertaken for the principals' sample which was too small for adequate multivariate regression analysis.

21 As a robustness check, the research team ran probit regressions with the same binary dependent variables. Results are very similar both in significance and magnitude of marginal effects. A further robustness check was considered by exploiting more information contained in the Likert-scale variables by running ordered probit regressions. They consider an ordinal dependent variable that takes the value of 1, 2 and 3 for (strongly) disagree, undecided and (strongly) agree, respectively. Since the dependent variable is constructed slightly differently than for the case of the LPM and probit estimations, the hypotheses tested are somewhat different and hence results are not fully comparable. Nevertheless, most of the implied tendencies remain true for the ordered probit estimations. These results are available upon request.

22 All regressors but age are binary variables.

Table 13. LPM Regressions Concerning Teacher Opinions on Seniority and Teacher Performance as Criteria for Pay

	Teacher promotion should be based on teacher performance	PKG should affect the teacher's salary	Teacher promotion should be based on seniority	Teacher salary should be linked to seniority
Female	-0.01	-0.05	0.03	0.10**
	(0.01)	(0.04)	(0.04)	(0.04)
Age	0.00	0.00	0.01***	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)
BA or higher	0.01	0.06	-0.10	-0.19**
	(0.02)	(0.07)	(0.08)	(0.09)
Passed cert.	-0.00	-0.07	-0.08	-0.07
	(0.02)	(0.06)	(0.06)	(0.05)
Civil servant	0.00	-0.01	-0.21***	-0.13***
	(0.02)	(0.05)	(0.05)	(0.05)
Public	0.01	-0.05	-0.00	0.03
	(0.02)	(0.06)	(0.07)	(0.06)
Semi-urban	0.00	-0.06	-0.14***	-0.13***
	(0.01)	(0.04)	(0.05)	(0.05)
% agree	97.2%	71.2%	34.2%	26%
Observations	500	500	500	500
R-squared	0.008	0.015	0.089	0.092

Note: Teacher sample. LPM regressions with binary dependent variable taking value of 1 if teachers agree or strongly agree with the statement and 0 otherwise. *, **, *** significant at the 0.1, 0.5 and 0.01 level. Standard errors are clustered at the school level in parenthesis. Columns are ordered from left to right in descending order after the share of teachers (strongly) agreeing with the corresponding statement.

Table 14. LPM Regressions Concerning Teacher Opinions on Indicators for Pay for Performance

	UKG should be part of teacher certification process	PKG should affect the teacher's salary	UKG should be linked to TPG	My promotion should partly depend on my students' test scores	If my students perform well in exams I should receive a bonus
Female	0.03	-0.05	-0.06	-0.04	-0.05
	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)
Age	-0.00	0.00	0.00	0.01**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BA or higher	0.02	0.06	-0.07	-0.05	-0.02
	(0.06)	(0.07)	(0.06)	(0.08)	(0.07)
Passed cert.	-0.05	-0.07	-0.21***	-0.02	-0.08
	(0.04)	(0.06)	(0.07)	(0.06)	(0.05)
Civil servant	-0.07	-0.01	-0.06	-0.01	-0.04
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Public	-0.05	-0.05	0.01	-0.04	-0.05
	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)
Semi-urban	-0.10**	-0.06	-0.11**	-0.13***	-0.12***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)
% agree	83.7%	71.2%	62.2%	61.6%	16.9%
Observations	427	500	427	500	500
R-squared	0.043	0.015	0.091	0.041	0.054

Note: LPM regressions with binary dependent variable taking value of 1 if teachers agree or strongly agree with the statement and 0 otherwise. *, **, *** significant at the 0.1, 0.5 and 0.01 level. Standard errors are clustered at the school level in parenthesis. Columns are ordered from left to right in descending order after the share of teachers (strongly) agreeing with the corresponding statement.

06 Conclusion

Discusses the opinions of principals, teachers, parents and students from **100** Indonesian schools

This paper discusses the opinions of principals, teachers, parents and students from 100 Indonesian schools concerning various issues related to teacher performance evaluation and PfP schemes. Multiple key insights are identified.

1. **First**, in general, the UKG and PKG are strongly supported by principals and teachers as teacher performance evaluators.
2. **Second**, PfP schemes involving the UKG and PKG are highly popular among principals and teachers.
3. **Third**, teachers strongly prefer PfP schemes over schemes based on seniority.
4. **Fourth**, while overall support is high among teachers, teachers in urban areas show a systematically higher level of support towards PfP schemes than teachers in semi-urban areas.
5. **Fifth**, teachers support the idea of student learning outcomes as a suitable indicator in a PfP setting.
6. **Sixth**, teachers and principals prefer PfP indicators that focus on teacher input, while parents favor teacher-parent and teacher-student interactions.

Finally, while the idea of education stakeholders (inclusive of school inspectors, principals, teachers, parents and students) as performance evaluators is generally supported, principals and teachers show stronger support for evaluators with a pedagogical background.

To shape the design and implementation of related policies and co-determine their success

This paper is informative for education policymakers. The attitudes of education stakeholders concerning performance evaluation presented in this paper are likely to shape the design and implementation of related policies and co-determine their success. By acknowledging the opinions of key education stakeholders, policymakers have the opportunity to contextualize appropriate policy design and minimize the risk of unintended effects. It should be noted, however, that opinion data, as presented here, has the inherent limitation of being subject to response biases related to social desirability or courtesy. However, the role of response biases in this paper is likely to be minimal since the majority of investigated survey items asked respondents to express opinions about potential future policies rather than rating past events.

Support PfP schemes.

In general, the analysis shows a clear general support of education stakeholders for PfP schemes.


ROSTER
BELAJAR KELOMPOK
- Nama kelompok
- Mata Pelajaran
- Hari/Tanggal
- Nama-nama anggota kelompok
1. GLODYSIA LODONG (LEDY)
2. FLAFIANUS TANGIS (FLAFI)
3. SRILIA JEMIA (SRI)
4. FORTUNATUS BARUNG (MBAKUPS)
5. KOLINTUS POGOL (KOLIN)
- Tugas Belajar Kelompok
- Mengerjakan PR
- Belajar mata Pelajaran hari ini
- Belajar mata Pelajaran BESOK
- Waktu belajar
ROSTER HARIAN
KELAS
-PKN
-BHSD
-AGAMA
KAMIS
IPS
BHSD
SBK
JUMAT
BHSD
IPA
SABTU
IPS
BELAJAR Kelompok
T/H Ajaran 2017/2018
Maret
April

References

Abadzi, Helen, and Stavri Llambiri. 2011. "Selective Teacher Attention in Lower-Income Countries: A Phenomenon Linked to Dropout and Illiteracy?" *Prospects* 41 (4): 491–506.

ACDP (Education Sector Analytical and Capacity Development Partnership). 2014. *Study on Teacher Absenteeism in Indonesia 2014.* Jakarta: ACDP

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

Bruns, Barbara, and Ben Ross Schneider. 2016. *Managing the Politics of Quality Reforms in Education Policy: Lessons from Global Experience.* Background Paper: The Learning Generation. New York: The Education Commission

Chang, Mae Chu, Sheldon Shaeffer, Samer Al-Sammarrai, Andrew B. Ragatz, Joppe de Ree, and Ritchie Stevenson. 2014. *Teacher Reform in Indonesia: The Role of Politics and Evidence in Policy Making*. Directions in Development: Human Development. Washington, DC: The World Bank.

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

Evans, David, and Anna Popova. 2015. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *World Bank Policy Research Working Paper* 7203. Washington, DC: The World Bank.

Gaduh, Arya, Menno Pradhan, Jan Priebe, and Dewi Susanti. "Impact evaluation of community empowerment and teacher pay for performance in Indonesia." Unpublished manuscript, last modified 15th December, 2018. Microsoft Word file.

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

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

Jinnai, Yusuke. 2016. "To Introduce or Not To Introduce Monetary Bonuses : The Cost of Repealing Teacher Incentives." *Economics & Management Series,* no. EMS-2016-08. Minamiuonuma: International University of Japan

Joshi, Anuradha. 2013. "Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery." *Development Policy Review* 31 (S1).

Kremer, M., C. Brannen, and R. Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340 (6130): 297–300.

Mann, H. B., and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18, 1 (1947): 50–60.

McKenzie, Phillip, Dita Nugroho, Clare Ozolins, Julie McMillan, Sudarmo Sumarto, Nina Toyamah, Vita Febriany, Robert J. Sodo, Luhur Bima, and Armand A. Sim. 2014. "Study on Teacher Absenteeism in Indonesia." Jakarta: Education Sector Analytical and Capacity Development Partnership (ACDP), Agency for Research and Development (Balitbang), Ministry of Education and Culture.

Ministry of Education and Culture. 2016. *Laporan Hasil UKG 2015* [Report on Results of the Teacher Competence Test 2015]. Jakarta: Ministry of Education & Culture, Directorate General for Teachers and Education Personnel.

Ministry of Finance, Indonesia. 2016. *Belanja Pemerintah Pusat, 2011-2016 (miliar rupiah). Jakarta.* http://www.anggaran.depkeu.go.id/dja/athumbs/apbn/2016Pendidikan.pdf

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011a. "Teacher Opinions on Performance Pay: Evidence from India." *Economics of Education Review* 30 (3): 394–403.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011b. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.

Murnane, Richard, and David Cohen. 1986. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review* 56 (1): 1–18.

OECD. 2016. *PISA 2015 Results (Volume I): Excellence and Equity in Education.* Paris: OECD Publishing.

Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, Armida Alisjahbana, and Rima Prama Artha. 2014. "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia." *American Economic Journal: Applied Economics* 6 (2): 105–26.

Sabarwal, Shwetlena, and Malek Abu-Jawdeh. 2017. "What Teachers Believe: Mental Models about Accountability and Absenteeism." Policy Research working paper; no. WPS 8454. Washington, D.C. : The World Bank.

UNCEN (Cendrawasih University), UNCEN, UNIPA, SMERU, BPS, and UNICEF. 2012. "'We Like Being Taught' A Study on Teacher Absenteeism in Papua and West Papua." Jayapura.

UNESCO. 2017. *Global Education Monitoring Report. Accountability in Education: Meeting Our Commitments. P*aris.

World Bank. 2010. *Transforming Indonesia's Teaching Force. Volume II: From Pre-Service Training to Retirement: Producing and Maintaining a High-Quality, Efficient, and Motivated Workforce.* Human Development Department, East Asia and Pacific Region. Jakarta: The World Bank

World Bank. 2015. *INDONESIA: Teacher Certification and Beyond. An Empirical Evaluation of the Teacher Certification Program and Education Quality Improvements in Indonesia.* Education Global Practice, East Asia and Pacific Region. Jakarta: The World Bank

World Bank. 2018a. Education Statistics (EdStats) Database. http://datatopics.worldbank.org/education/.

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

Appendix

Table A1. Survey Locations

Province	Name of City/District	Target/Neighboring	Geography
Jawa Barat	Kota Banjar	Neighboring	Developed
	Kota Tasikmalaya	Neighboring	Developing
Bali	Kota Denpasar	Neighboring	Developed
Nusa Tenggara Barat	Kabupaten Dompu	Target	Developing
	Kota Bima	Target	Developed
Nusa Tenggara Timur	Kabupaten Manggarai Timur	Target	Very remote
	Kabupaten Sumba Barat Daya	Target	Remote
	Kabupaten Sumba Barat	Neighboring	Remote
Sulawesi Utara	Kota Bitung	Target	Developed
	Kota Manado	Neighboring	Developed

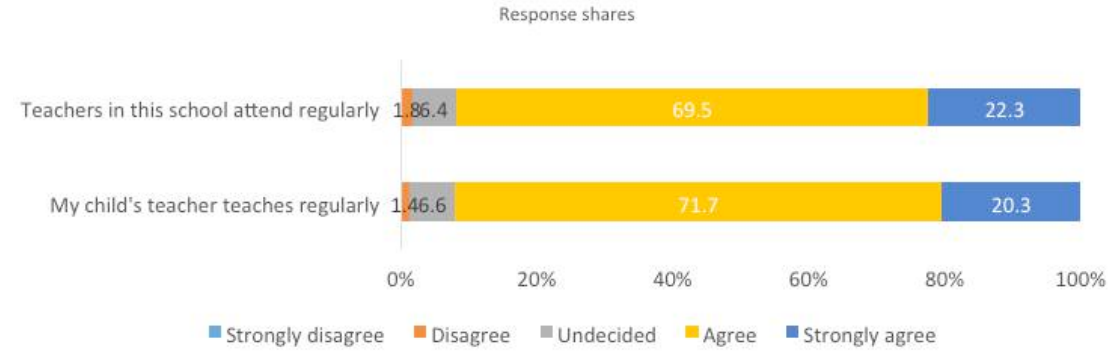
Note: During the first stage of research five cities-districts were purposively selected to represent heterogeneity in terms of geography (i.e., target city-district). In the second stage, for each of the five cities-districts one neighboring city-district was also selected (i.e., neighboring city/district).

Table A2. Extended PKG List of Teacher Competencies for Teacher Performance Evaluation

Can assess students' characteristics
Master educative teaching and learning theory and principles
Can develop curriculum into lesson plans
Conduct teaching and learning activities
Develop potential of student
Improve learning outcomes*
Improve average learning outcome at school*
Improve communication with students
Can assess and evaluate students
Behave in line with moral, social, cultural and religious norms
Are role models
Have strong work ethic, sense of responsibility, and sense of professional pride
Are tolerant and non-discriminatory
Able to communicate with teachers, parents, educational personnel, students and community
Are able to motivate parents*
Master their subject matter
Continuously improve their teaching competence, knowledge, and skills

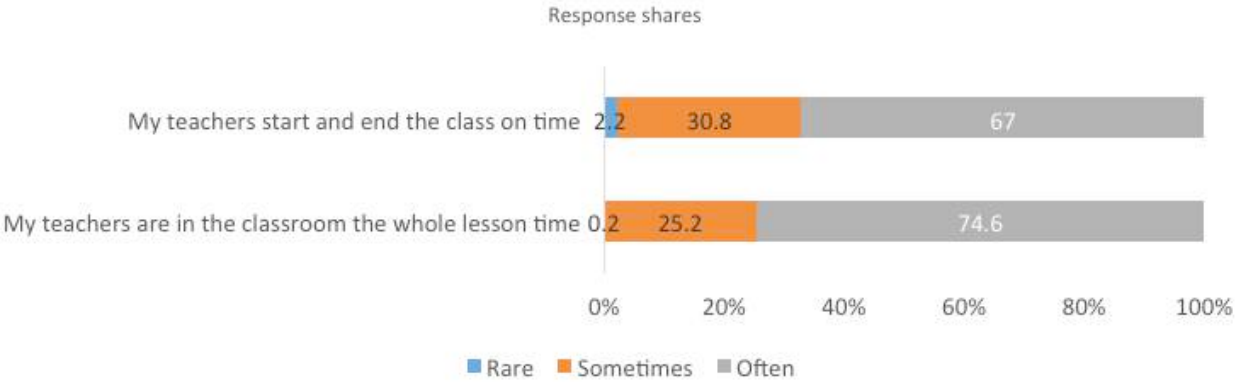
Note: * Three additional competencies that were added to the PKG list with respect to student learning outcomes.

Figure A1. Parents: Teacher Absenteeism



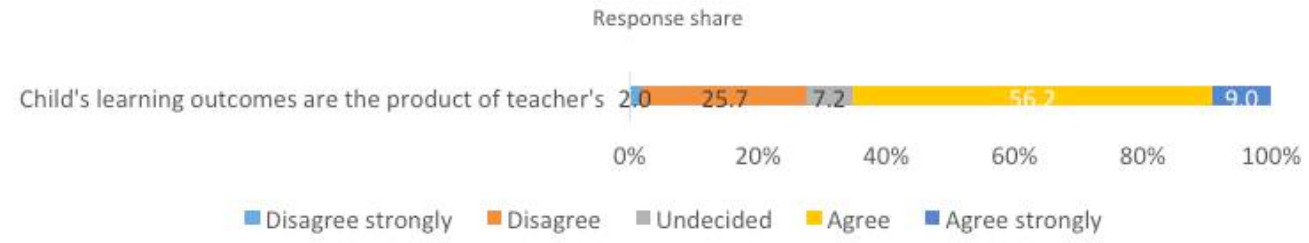
Note: 502 observations. Source: KIAT GURU Urban Opinion Survey 2017.

Figure A2. Students: Teacher Absenteeism



Note: 500 observations.

Figure A3. Parents: Teacher Ability and Child's Learning Outcomes



Note: 502 observations.