ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics & Business Economics

The Effects of Performance Pay for Indonesian Rural Teachers: Multitask Moral Hazard

Student name: Nityananda Risqika Andjani

Student ID number: 523536

Supervisor: Prof. Robert Dur

Second assessor: Mareen Bastiaans

Date final version: 1 August 2022

Table of content

Abstract

This paper presents results regarding the effect of monetary incentives for teachers based on an experiment among rural teachers in Indonesia. Teachers were rewarded with cash bonuses for improving their students' performance in primary school exams or through teachers' attendance to class. One of the biggest challenges of a performance-pay incentive is the potential of the multitasking moral hazard, including potential trade-off between two roles teachers have in school. It is important to analyse the benefits and losses that arose through the introduction of performance-pay incentives. Using data from an experiment across rural areas in Indonesia, potential moral hazard from the experiment will be evaluated. The data will be analysed through a difference-in-difference model, using time spent on teaching and time spent on secondary roles reported by teachers in different treatment groups. The results show that through the incentive based on student test scores, teachers increase hours spent on teaching tasks and reduce hours spent on secondary roles. Whereas when teachers are incentivized through teacher attendance, teachers are more likely to reduce time spent on secondary roles whilst leaving teaching hours unaffected. Findings shall be taken into account to ensure desired outcomes are not at the cost of the school, and shall then be evaluated to further analyse the optimal design for performance-pay wage contracts.

---

1. Introduction

Improving the quality and effectiveness in public education has always been a global discussion. Especially in developing countries where the challenges that arise are always more prominent. In Indonesia specifically, even though the government spends more than 20% of their total government budget on the education sector (Afkar, Luque, Nomura, & Marshall, 2018), they are ranked amongst countries with mean education performance below the OECD average in terms of student performances in science, reading and mathematics (Programme for International Student Assessment, 2015). Apart from that, the national early childhood development enrolment ratio shows high inequality between different regions, where urban areas perform better than rural ones (Unicef, 2017).

The Indonesian education budget is dominated by large fixed costs for salaries of teachers, in expense of non-salary budgets for quality improvement activities such as training and other operational costs that are seemingly as important to improve student outcomes (Afkar, Luque, Nomura, & Marshall, 2018).

Empirically, substantial increases in spending in rural public schools does not directly provide effective improvement in education quality (Shikalepo, 2020). Considering the high scarcity of resources in a developing country like Indonesia, policymakers are always debating in regards to the effectiveness of different policies and programs to improve performance including a performance pay model for their teachers. Looking at the large fixed costs of salary for teachers, governments across the world experiment with formal incentives, specifically performance pay. Performance pay is often debated to increase productivity of workers considering all individuals are rational, and is deemed cost-effective in terms of the outputs provided. For example, performance pay increases productivity to up to 36% of output of low-skilled workers that assemble glass as well as reducing absenteeism (Lazear, 2000). However, the external validity of such wage contracts are often questioned when applied in other industries or jobs that do not have measurable output easily, including teachers.

Performance pay has been a successful wage model implemented in various industries, however, cases within the education public sector still have mixed results. Teachers are also more hostile towards performance pay based on student test scores, leading to long-term unforeseeable impacts. Especially in developing countries, such measures are now often implemented in hopes of increasing productivity of teachers. However, there are some potential negative consequences that are questioning the effectiveness of such programs.

Various programs to improve school quality in Indonesia have been done by providing more resources such as grants through the *Bantuan Operasional Sekolah* (BOS) program (Al-Samarrai, 2014). This incentive provided schools with more resources, however, it does not seem to have significantly improved student performance and since the budget is allocated internally, it is difficult to measure how it affects quality of education. With more research on teachers being compensated based on their performance, there has been growing interest in directly measuring and rewarding schools and teachers on the basis of student learning outcomes. The same program has been evaluated before using the same data by the World Bank (2021).

Performance pay does induce a short-run beneficial effect towards the intended targets, however, does not entirely capture the long-term performance (Leigh, 2012). One of the most debated challenges from performance pay is the multitasking moral hazard. This arises through a performance pay measure where teachers might be incentivized on certain tasks at the cost of other

activities they need to fulfil (Mbiti, Muralidharan, Romero, Schipper, Manda, & Rajani, 2019). In the education space for teachers in rural areas, they often play more than one role apart from teaching. This includes tasks around organisation of the school itself such as becoming tutors, running extracurricular activities, parent guide, administrative person, community engagement, and many more. Thus, due to the lack of workforce within each school, teachers are also often multitasking between administration and teaching activities (World Bank, 2008).

Given the centrality of this question to labour and personnel economics, previous empirical literature has attempted to study the impact of conditional and unconditional pay raises on workers' effort and productivity, with varying results. Empirically, it is evident that a performance-based pay significantly resulted in higher student performance in India (Muralidharan & Sundararaman, 2011). Similarly, Lavy (2009) also showed that it leads to significant improvements in student performance, measured through test rates, pass rates, and mean test scores in Israel. Similarly in Tanzania, it is found that a performance-based pay for teachers significantly improved student performance compared to grants (Mbiti, Muralidharan, Romero, Schipper, Manda, & Rajani, 2019). Ree, Muralidharan, Pradhan, & Rogers (2018) also showed that an unconditional pay increase is not an effective policy to improve the effort and productivity of public-sector settings. This led to an ongoing research to see the effectiveness of such wage design. The big challenge of this moral hazard should be evaluated across a broader range of firms and industries, with a particular focus on workers who have to exercise discretion in their work such as teachers (Bartel, 2017). Some characteristics in discretion is that it allows creativity within the work. Bartel (2017) also calls these kinds of employees 'knowledge workers'.

When a performance pay measure is implemented, teachers could focus on the performance measure rather than the job itself, resulting in a multitasking moral hazard. With student test scores and teachers' attendance in the classroom as a performance measure, the paper will question, *"Does performance pay cause a multitasking moral hazard within teachers in rural areas of Indonesia?"*

Through a difference-in-difference model, this paper will compare the hours of time spent on teaching and on secondary roles that are based upon the operations of the school itself. Further, this paper analyses whether the effectiveness of the program is at the cost of other tasks. As it is difficult to realistically measure performance on the supporting task, this research will use teachers' hours spent on certain tasks as a measurement for the investigated variable. This is because all the supporting tasks of each teacher differ, from becoming librarians, head of administration, teaching

different extracurricular activities and parent-teacher coordinator, and many others. Using the proxy of hours spent, the model will compare the hours spent between treated and non-treated groups of performance pay, as a measure of effort of the teacher on the supporting task.

To discuss the findings in this paper, it shows significant positive results towards language and mathematics test scores but negative insignificant effects towards teachers attendance show that the effectiveness of performance pay towards the two performance measures is mixed. This aligns with previous literature (Leigh, 2012) that argues that it is impossible to find the optimal performance pay design that best reaches the desired outcome. With the question of the multitasking moral hazard, it is a possibility that teachers with secondary roles would be incentivized to spend more effort and time in teaching related activities, at the cost of their secondary role tasks.

It is found that through the incentive based on teacher attendance, teachers significantly reduce hours spent on secondary roles by 0.573 hours. This result adds additional costs from the performance pay scheme, especially understanding that using teachers attendance as a performance measure did not significantly result in increasing attendance of teachers itself. However, it also leaves teaching hours unaffected. When teacher incentives are through student test scores, it is statistically significant that they are more likely to increase hours spent on teaching tasks by 0.310 hours, whereas it is also more likely to reduce hours spent on teaching tasks by 0.398 hours. Therefore, there is strong evidence for multitasking in this data. Both results are statistically significant. With both treatment groups, it can also be seen that teachers' division of time spent is also affected by the teacher's age and certification. The older the teacher's age, the more likely performance pay is to decrease the time spent on secondary roles and reduce time spent on teaching activities. On the contrary, for certificated teachers, they are more likely to increase time spent on teaching roles and reduce time spent on secondary roles in response to the monetary incentive.

The same program has been evaluated before using the same data by the World Bank (2021). They conclude that the program showed mixed results, with larger positive impacts towards student test scores compared to teacher attendance. They also concluded that there were still potential long-term challenges from this program. My analysis differs in the specificity of the performance measure by taking the raw scores and attendance, and focuses on the impacts of performance pay towards multitasking problems. My main contribution as compared to the World Bank (2021) thus is that even though there are positive effects towards student performance, there is evidence of multitasking problems where teachers would invest less time on other activities than teaching.

In line with previous literature, the results of this study are rather mixed in terms of the effectiveness of performance pay. Nevertheless, it provides evidence on the multitasking problem. Teachers spend more time on verifiable and incentivized tasks compared to the secondary tasks. Especially in Indonesia, as the secondary tasks are beneficial activities for the schools, it is a challenge in the long-run. While this research is set in the context of schools and teachers, this paper also contributes to the broader literature on performance pay in organisations in general and public organisations. There are still large gaps in multitasking, as it is important to see the impact of multitasking including in different sectors such as the public sector (Bartel, 2017).

2. Theoretical Framework

As most performance-based pay for teachers are based upon students' performance measured through test results (Muralidharan & Sundararaman, 2011; Lavy, 2009; Mbti et al, 2019), a potential challenge is that it provides a high incentive for teachers to conduct unethical behaviour to raise grades. Such actions may include cheating, focusing on drilling, or making the exam questions easier. In Israel, it is evident that the improvement of maths and English student outcome improvements were mediated through changes in teaching methods, enhanced after-school teaching, and increased responsiveness to students' needs (Lavy, 2009).

The mechanism of using student test results as a performance measure for teachers also does not capture the improvement in non-test-based results. These may include development in students' personal attitude or other skills. Previous empirical results also emphasise that performance pay discourages teachers from working together, as teachers are then more focused on personal gains (Podgursky & Springer, 2007). Apart from that, it is also more inclined to certain unfairness as certain performance measures might be influenced by random external factors. For example, if an intervention focused on student test scores, it would also be influenced by personal, family, neighbourhood and community inputs, or other idiosyncratic shocks. Performance measures like these are then very highly defined with uncertainty, which risks the hypothesis that a multitasking risk increases (Foss & Laursen, 2005). If teachers are risk averse and schools are risk neutral, it can cause lower sensitivity for teachers to then be incentivized towards the desired outcome (Balmaceda, 2009).

The performance-based wage contract still has many mixed results in improving students and teachers performance globally. Some saw increases in student test scores mainly around mathematics and language in other developing countries, such as India (Muralidaran &

Sundararaman, 2006), but some saw insignificant or negative impacts towards student performance in Rwanda (Glewwe, Kremer, Moulin & Zitzewitz, 2004). The mixed results can also be found specifically for the public sector (Weibel, Rost, & Osterloh, 2010).

Baker (1992, 2002) argues that these wage contracting models may increase the likelihood of undesired outcomes such as the multitask moral hazard problem (Holmstrom & Milgrom, 1991). Andrabi & Brown (2022) also saw a multitasking problem in schools when performance incentives are applied through subjective (manager evaluation) and objective schemes (student test scores). They found that student test scores increase at the expense of quality of teaching, evident through the reduction of non-test score student outcomes as well as reduction in teaching effort.

When teachers are located in rural areas, they are more likely to have much more responsibilities due to lack of human resources. Thus, most teachers in rural areas of Indonesia would have secondary roles at school, such as management, administrative, community-building or extracurricular activities. All the secondary roles are also arguably important for the operations of the school itself, which relates to the quality of education and potential student outcomes. Previous papers also show how student achievement is highly dependent on the quality of such activities, including institutional operations, curriculum building and teacher's education (Woessmann, 2016). This raises the challenge of whether the performance-based pay would effectively improve certain performance measures, at the cost of effort on these secondary roles or activities.

To find the most optimal model, Lavy (2007) states that the performance pay incentive schemes should align performance with ultimate outcomes and must be monitored closely, hence this improves teacher performance rather than only student test scores. With student test scores as a performance measure, it may also cause a multitasking problem for teachers. For example, Task 1 represents teaching and Task 2 represents activities designed to increase scores on exams. As there are various ways teachers may take to reach higher student test scores, activities to increase scores may be caused by unethical behaviours such as drilling, coaching on items likely to be on the test, and perhaps cheating (Muralidharan & Sundararaman, 2011). Some literatures also saw similar negative multitasking distortions in previous empirics. For example, Jones, Tonin & Vlassopoulos (2018) saw pay-for-performance led to negative effects on the non-incentivized dimension or second task. Apart from that, a successful pay-for-performance intervention in Kenya saw that basic work was unaffected, but test-related activities such as test drilling sessions increased (Glewwe, Nauman, & Michael, 2010), similar to a condition in the private sector (Hellmann & Thiele, 2011). In Hellman &

Thiele (2011), it is found that performance pay contracts only work for tasks that are 'measurable ex ante', or activities that are easily measurable which is not the case for teaching. Teaching also requires activities for over a longer period of time with no clear measurable point, thus, theoretically would not be suitable for the performance pay scheme.

The formal modelling of the multitasking moral hazard is based upon Holmstrom and Milgrom (1987) linear principal-agent model, which assumes a linear performance incentive. The model also highlights conditions and implications of how tasks get allocated to different jobs (Holmstrom & Milgrom, 1991) and the multitasking problem is elaborated by Baker (2002) and Neal (2010) which assumes that the agent's costs depend only on the total effort the agent puts into all of the agent's tasks. Thus, it will show that an increase in an agent's compensation in any one task will cause reallocation of effort away from other tasks. The model also assumes that the agent's effort is a homogenous input that can be allocated among the tasks however the agent likes. This is quite aligned with most performance-based incentives on teachers as a public worker. For simplification, the model presented assumes there is no risk aversion on the model.

In the model, we assume that teachers are the agents in question. As part as their job, they have to engage in two types of tasks, T_1 and T_2 , where T_1 represents the primary job, specifically teaching and T_2 represents the activities related to increasing test scores such as drilling, coaching and cheating. Assume that e_1 and e_2 represent the total effort exerted on the two tasks.

Holmstrom (2016) is based upon the general linear multitasking model. It is assumed that an agent's effort is a homogenous input that can be allocated between tasks and that effort in the two tasks are perfectly substitutable in the agent's cost function. The two tasks have different measures. The first task can be perfectly measured, such as student test scores. The second task however, is harder to measure, such as the performance in secondary roles.

Holmstrom and Milgron (1991) starts by modelling that the performance measure by

$$P = g_1 e_1$$

where g_1 are the marginal products of total effort on teaching related activities. For simplicity, we assume that $g_1 = 1$. Through this measure, the contractor creates a wage contract. The wage contract is created as a function of P:

$$w = s + b \cdot P$$

where w is the total wage, s is the salary and b is the bonus rate paid per unit of P. Following Holmstrom & Milgrom (1991), the optimal bonus rate b^* depends on the functional form of the cost function. The teachers utility function is modelled by:

$$U = w - C(e_1, e_2)$$

where w is the wage contract and $C(e_1, e_2)$ is the cost associated with any combination of T_1 and T_2

Let the principal's objective be

$$B(e_1, e_2) = p_1 e_1 + p_2 e_2$$

where p_1 and p_2 measure the value of teaching and secondary roles.

Let the agent's cost function is

$$C(e_1, e_2) = c(e_1) + c(e_2) + c(e_1 + e_2)$$

Agent chooses effort level for each task $e = (e_1, e_2)$. To solve the model, assume that the cost of effort for agent is $C = \frac{1}{2}\theta^2$. The agent's cost function would then be modelled as

$$C(e_1, e_2) = \frac{1}{2}\theta {e_1}^2 + \frac{1}{2}\theta {e_2}^2 + \frac{1}{2}\theta (e_1 + e_2)^2$$

where the first two components measure the cost of effort for teaching and secondary roles respectively. The last component measures the dependency of the two effort choices, as if teachers already spend more effort on a certain task, it makes it more costly to exert more effort on secondary roles.

The utility function of the agent is then modelled as $U = w + \gamma\rho - C(e_1, e_2)$ where w is the wage contract, γ is the intrinsic motivation teachers have, and ρ is the production. ρ is modelled by $\rho = k(e_1 + e_2)$. When maximising the agent's utility function with respect to e_1 and e_2. This will then find that the first-order-conditions is as follows:

$$\gamma k + b - \theta e_1 - \theta(e_1 + e_2) = 0 \qquad (1)$$

$$\gamma k - \theta e_2 - \theta(e_1 + e_2) = 0 \qquad (2)$$

From the two first-order conditions, we can see that $\theta e_i + \theta(e_1 + e_2)$ measures the marginal costs of effort, whereas b measures the additional financial bonus for the agent from exerting more effort on teaching. In this case, the agent would choose to exert more effort for teaching. With this, it would also risk increasing the overall marginal cost of effort in performing any tasks, as well as increasing the marginal cost of effort on secondary roles. When solving the first-order conditions

above, we find that $e_1 = \frac{\gamma k+2b}{3\theta}$ and $e_2 = \frac{\gamma k-b}{3\theta}$. We can see that the bonus exerted on the two tasks differently.

In general, if $C'(0) < 0$, the agent will always choose a positive level of total effort which is an important success of the performance pay measure. However, the multitasking problem would still occur, transferring effort from secondary roles to teaching. In this case, lowering the bonus incentive is advantageous for the increase in value from secondary roles. However, as secondary roles are unverifiable, it would not be effective. As the two tasks are perfect substitutes and attention to these secondary roles are still valuable for the principal, any incentive on teaching will have to be paired with a strong incentive on secondary roles.

If the total hours spent on working does not increase, it can be assumed that the performance pay incentive may lead to activities such as cheating as teachers do not put in extra hours to provide higher quality teaching, but instead focus on the hours spent on improving student test scores (Dixit, 2002). In most cases, this quality is always unverifiable. If total hours of effort improve and it does significantly reduce the hours spent on secondary tasks, it may seem that a multitasking moral hazard does not significantly impact the effectiveness of the program.

It can be argued that with the basis of personnel economics, the indirect impact of this intervention may indirectly offset the beneficial impacts of the performance pay policy itself (Weibel, Rost, & Osterloh, 2010). It is likely performance pay will cause a multitasking moral hazard which allows teachers to put more effort into tasks that are aligned with performance measure, at the expense of effort on their secondary roles or activities that is evidently also important to the school organisation. Hence, the effectiveness of the chosen performance measure could be reevaluated. It is valuable to evaluate this to ensure alignment between the government as the principal, and teachers as agents within a principal-agent theory model.

3. Data

3. 1 Data Context

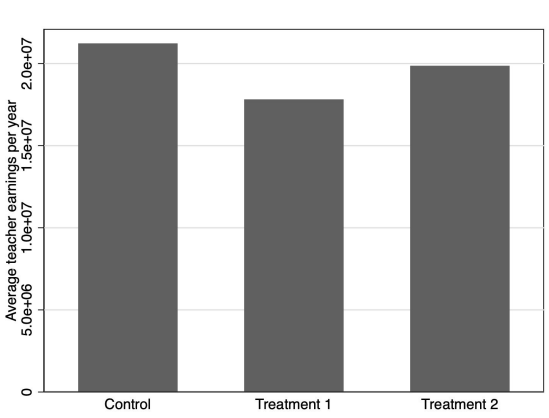
The Indonesian Ministry of Education and Culture (MoEC) and the Indonesian National Team for the Acceleration of Poverty Reduction (TNP2K) and in collaboration with the World Bank, established a randomised control trial large-scale experiment through the “KIAT” project, aiming to improve education quality within Indonesia’s rural areas. They conducted interventions of social accountability mechanisms and performance pay for teachers which is modelled similarly to

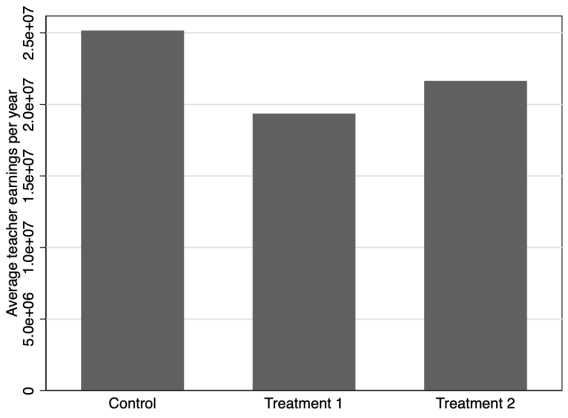
Holmstrom & Milgrom (1991) principal-agent model with a fixed salary and bonus depending on performance that is based on the fixed-salary. Thus, it could also induce a negative bonus for poor performance.

The intervention was implemented with two different types of performance measures: teachers attendance and student test scores. Teachers attendance was measured through an internal attendance form, as well as a monitoring system with installed cameras to track working hours. The attendance is dependent on the daily percentage attendance. Teachers whose total cut exceeded 15 percent of their monthly attendance will lose their monthly bonus. The design allows negative incentives for poor teacher performance.

For the incentive based on student test scores, it was measured through a final exam for the Indonesian language and mathematics subject. The performance pay contract was designed with a regular civil servant pay and additional bonuses for performance. This is also aligned with the model presented by Holmstrom and Milgrom (1991) where the wage contract is modelled with a fixed salary and a bonus rate depending on a scorecard of their performance. There is a maximum of the bonus which is standardised throughout the program. The amount of bonus received will then be based on the average score of the students. For example, if a teacher received an 80 percent score, then the teacher would receive 80 percent of the bonus allowance.

The bonus allowances in all groups were paid on a quarterly basis for all treatment groups. Prior to the experiment, on average, non-certified teachers had an average monthly income of 4.6 million Rupiah, whereas certified teachers had an average monthly income 8.4 million rupiah. In the total sample, approximately 66 percent were not certified. Those that are not certified, would often only gain a monthly income averaging 0.55 million rupiah.





A B

Figure 1. (A) Average teacher earnings in pre-treatment period (2016); (B) Average teacher earnings in post-treatment period (2018)

The average teacher earnings in the pre-treatment period for the whole sample is 16.7 million rupiah per year, which is roughly only 1,4 million rupiah a month. This is significantly lower than the average monthly income of Indonesia of 2,556 million rupiah. This aligns with the context of the data where they chose the rural and poorest regions in Indonesia. As teachers are not paid well prior to the intervention, teachers would become more incentivized through a performance-based pay. This would allow teachers to gain more income whilst focusing on work at school. The low pay may also highlight that teachers exhibit loss aversion. In this case, it would be more effective to cut existing allowance from the government rather than additional bonus from the basic one (Fryer, JR., Levitt, List, Sadoff, 2018).

As randomization took place at the region level, the difference in average salary happens by design as the schools selected for the treatment groups are from different regions of the country. Each region would have a different average salary, and also the context of the school. Either way, the difference is not substantial considering that it is still within the ranges of salary around rural regions. However, when using a balance test, it also found a significant difference between the means of pre-treatment with a p-value of 0.001. This would mean that there are significant differences between the groups that might be caused by the randomization that happened at a region level. Comparing the two periods, Figure 1(B) shows that, the average income increases for all three groups. It is logical as the project raised the budget for teacher salary through the bonus allowance allocated to teachers.

Between the treatment and control group, Figure 1 also shows that the treatment group based on teachers' attendance have the lowest average income per year whereas the control group has the highest average income per year. This might also influence the later results of the research as those in treatment would have higher incentive to exert higher effort and earn more money through merit-based pay.

The experiment starts with a baseline survey in 2016 conducting all prior information from demographics, characteristics, work satisfaction, day-to-day work activities, hours spent and income. When the intervention started, both treatment and control teacher groups were informed of the

new wage contract models with bonuses measured through different incentives. The intervention lasted for 2 years starting from 2017. The first treatment group based on teachers' attendance would have to sign their attendance through monthly administration, as well as monitoring using newly installed cameras in classrooms. This might increase the cost of effort for teachers as they would have to spend a bit more time on administrative work. For the treatment group on student test scores, teachers would then understand that they will get a bonus based on their student's performance, without any additional changes to their jobs. The student tests are measured through two subjects, Indonesian language and mathematics. The tests were conducted in 2018, as well as in 2019 as a follow-up survey.

3.2 Data Source

The panel data is obtained from 4 surveys: baseline (2016), attendance monitoring data (2016-2017), endline (2018), follow-up survey (2019). The baseline and endline survey involved 270 primary schools spread in 235 remote villages across 10 sub-districts selected in Sumatra-Java, Nusa Tenggara, Kalimantan, Sulawesi and Maluku-Papua based on the highest teacher absenteeism. The follow-up survey only involved 176 primary schools. All of the surveys were conducted through a face-to-face interview. In the attendance monitoring data, it also includes data from the camera monitoring of teachers attendance at school. The average rural primary school enrollment is usually small (Muralidharan & Sundararaman, 2011), and in this study, the total enrollment of each school is approximately 50 students.

In the program evaluation report the World Bank (2021) measures the performance of teachers by a scorecard that combines the quantitative measurement of attendance and student test scores, but also includes a score from the community enhancement activity. However, in this paper the replication will focus on the effectiveness of the performance measure itself. Thus, the analysis will take into account the raw data of the two different performance measures. For teacher attendance, the total attendance percentage in school per month will be used. Whereas for student test scores, we will take into account the raw data for Indonesian language, mathematics and average test scores of each student. The data also collects individual information of teachers' demographics (age, and gender) and job characteristics (certificate status) that will be taken into account. In the analysis, we will take into account the certificate status of teachers which may impact the time needed to spend on teaching compared to others.

To determine the time allocation between the two job roles, we will look at the total hours attending school per day compared to the total hours teaching per day. In this design, the study focuses on teachers that have secondary jobs at the school. As aligned with previous literature, a large proportion of 78.24% of teachers in the sample have secondary roles (Appendix 1).

There are various secondary roles that are vital for the running of the school. Secondary roles here include operational, administrative or management roles, extracurricular activities, management activities such as teacher or parent coordinators, and many more (Appendix 2). The performance of teachers in their secondary roles are hard to track. Unlike classroom activities, the secondary roles that are available at these schools are different per individual teacher, the nature of the work is usually very different and all of the roles have varying performance measures (Appendix 2). To measure hours spent on secondary job roles at the school, a new variable is created by subtracting the total hours working to the total hours spent on teaching or exam related activities.

Nevertheless, there are some limitations to the data and implemented experimental design. Firstly, a limitation would be that the sample is non-random as it focuses on rural areas with significantly lower economic development. Rural areas also usually need a more localised policy (OECD, 2019), making the context found in this data lacking in exogeneity when looking at the multitask moral hazard from performance pay incentives. Apart from that, the face-to-face survey format also indicates that most information only takes into account the self-perceived hours of time spent on different activities including teaching that relates with the performance measure of student test scores and other administrative or extracurricular activities.

In terms of the design of the impact evaluation research (World Bank, 2021), the teacher attendance might not be the best performance measure as it could cause the Hawthorne effect. This effect allows teachers to perform better throughout the monitoring period. For example, activities like face-to-face interviews and usage of cameras as monitoring technologies become a signal to teachers that they’re part of the experiment with the government. If the experiment is no longer applied, teachers' behaviours could change in the longer-run especially with less monitoring.
The chosen measure for student performance of test scores is also not that suited for the age group in the sample. All the schools are primary schools with students aged 6-11 years old. Many papers show that students in primary school have various ways of developing and learning, including other cognitive or personal development not measured in the tests that is argued to be formed during the years of elementary school (Wood & Kaszubowski, 2008). However, no test can capture both

academic and non-academic performance. With test scores only measured through Bahasa Indonesia and mathematics, teachers would be incentivized to only focus on a narrow subset of activities that focuses on the year end tests (Hoxby, 2002).

3.3 Empirical Specification

To check the evidence of the effect of the intervention from the previous program evaluation results (World Bank, 2021), we will conduct a difference-in-difference method to see the effect of performance pay on teachers' attendance and student test scores.

$$y_{it} = \alpha + \rho T_i + \gamma t + \beta(T_i \cdot t) + \epsilon_{it} \qquad (3)$$

Where the dependent variable is y for teacher *i* at time *t,* looking at three different variables including 1) teacher attendance; 2) Indonesian language student test scores; and 3) mathematics language student test scores. Model (3) will provide a general result and scale of the impacts of the performance pay towards the performance measures.

To see whether there is evidence of multitasking moral hazard within this intervention, a difference-in-difference method will be implemented by comparing the time allocated to teaching and time allocated to secondary roles, which is modelled by:

$$y_{it} = \alpha + \rho T_i + \gamma t + \beta(T_i \cdot t) + \epsilon_{it} \qquad (4)$$

Where the dependent variable y is hours spent for teacher *i* at time *t,* looking at two different variables including 1) hours spent on teaching and exam related tasks; 2) hours spent on secondary role tasks.

For both models (3) and (4), *t* is a dummy variable of time where t is the time of pre-treatment and post-treatment period. *T* is the dummy of the treatment group where T = 0 if the teacher is in the non-treatment group, and T = 1 if the teacher is in the treatment group. For time, t = 0 is the pre-treatment period whereas t = 1 and t=2 is the post-treatment period. The interaction term, $(T_t \cdot t)$ multiplies the treatment group and the treatment period to measure the average differential change in *y* from the first to the second time period of the treatment group relative to the control group. Thus, the treatment effect will be captured by β.

Model (4) will allow the paper to estimate the effect of the performance pay intervention by comparing the changes in outcomes over time from 2016-2019, between the treatment and control group of schools participating. Since the intervention was made at an aggregate level, a difference-in-difference model is preferred.

3.4 Descriptive statistics

Table 1.

Variable Descriptions

Variable	Description
Treatment	Dummy variable of intervention group (0 = control, 1 = treatment)
Time	Dummy variable of pre-treatment and post-treatment period (t = 0 = pre-treatment, t= 1, 2 is post-treatment)
Female teacher	Gender of teacher (0 = male, 1 = female)
Teachers Age	Age of teacher (years)
Certificated	Dummy variable of whether the teacher is certificated (0 = not certified, 1 = certified)
Teacher Attendance	Teacher attendance per month by percentage (%) of attendance over total number of days a month, monitored through monitoring camera app throughout monitoring period (2016-2018)
Bahasa Indonesia student test scores	The Bahasa Indonesia student test scores of students taught per teacher on a scale of 0-100.
Mathematics student test scores	The mathematics student test scores of students taught per teacher on a scale of 0-100.
Hours spent on teaching	Total hours spent on teaching-related activities per week
Hours spent on secondary role	Total hours spent on other secondary activities per week
Total work hours	Total hours spent working at school (sum of hours spent on teaching and hours spent on secondary role)

Note: The data is obtained from The World Bank Database (Improving Teacher Performance and Accountability (KIAT Guru) Impact Evaluation, Survey, 2016-2019). This is not the comprehensive list of all variables for which data was collected.

Variables used are described in Table 1, which will be used to identify the effectiveness of the performance pay program as well as seeing if a multitasking moral hazard occurs. The model will also take into account certain individual teacher characteristics such as age, gender, and certification statuses. Many papers show a negative relationship between age and productivity. Similarly in the education landscape, age has a negative effect towards general teaching effectiveness (Homer, Murray, & Rushton, 1989). Older teachers may spend more time on work that requires less effort, such as administrative work in their secondary roles. When looking at teachers' gender in rural areas, this would also highly influence teacher's choice in hours spent on the two tasks. With a wide gender gap in Indonesia's rural areas (FAO and UN Women, 2021), the patriarchal culture would disincentivize women to work longer hours, or conduct more work similar to men due to discrimination or social norms. Lastly, certification statuses would impact the level of individual cost to perform or exert effort in teaching. Past papers saw that in developed countries, teachers with higher qualifications such as a master's degree do not necessarily have an added value (Hanushek & Rivkin, 2004). However, the governmental policy of the certificated teachers in Indonesia ensures legally a higher pay. This would potentially reduce their hours of work compared to teachers who are not certificated, however, due to intrinsic motivation, they are more likely to also spend more time on teaching related activities than administrative work.

However, the model does not take into account students' characteristics. This assumes that multitasking moral hazard only occurs due to the teacher's cost of effort. However, a potential addition to strengthen the model is a school-level control characteristics including variables such as the total number of teachers and other measurable school performance such as financial or accreditation.

Table 2.

Summary of teachers and students characteristics

Variables	Control (1)	Bonus incentive: Teacher Attendance (2)	Bonus incentive: Student test scores (3)	All Sample (4)
Teacher age	42.308	43.069	41.277	41.863

Female teacher	0.166	0.130	0.176	0.168
Teacher certificated	0.218	0.243	0.247	0.233
Teaching hours				
Baseline teacher hours spent on teaching per week	15.309	15.373	14.167	14.427
Endline teacher hours spent on teaching per week	15.549	15.817	14.681	14.964
Baseline teacher hours spent on secondary role per week	8.853	9.022	9.986	9.775
Endline teacher hours spent on secondary role week	9.2	9.386	9.817	9.676
Baseline teacher total hours of work per week	24.467	25.329	23.809	24.047
Endline teacher total hours of work per week	24.5	26.345	24.089	24.35
Incentives				
Baseline teachers attendance per month (%)	0.757	0 .771		
Endline teachers attendance per month (%)	0.758	0.785		
Baseline Indonesian language test scores (0-100)	35.567		39.252	
Endline Indonesian language test scores (0-100)	44.740		65.212	
Baseline mathematics test scores (0-100)	35.579		38.123	
Endline mathematics test scores (0-100)	46.123		56.80	
Number of observations	16,080	6,797	15,368	

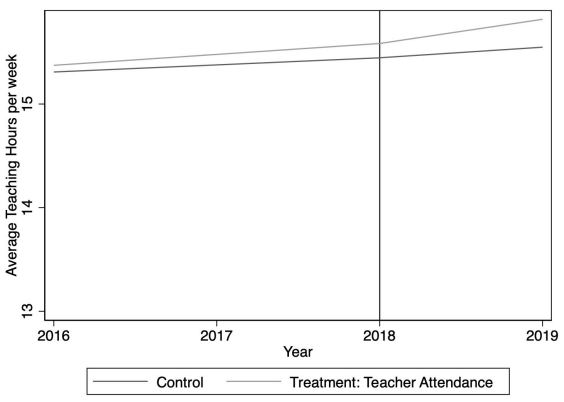
Note: The data is obtained from the World Bank Database (Improving Teacher Performance and Accountability (KIAT Guru) Impact Evaluation, Survey, 2016-2019).

The data was composed from a total number of 233 schools, 1,422 teachers and 23,488 students (Appendix 3) , an approximately balanced sample, with no significant differences between the groups (Table 2). Table 2 also shows that the basic mean teacher's age for all samples is 41.863, with

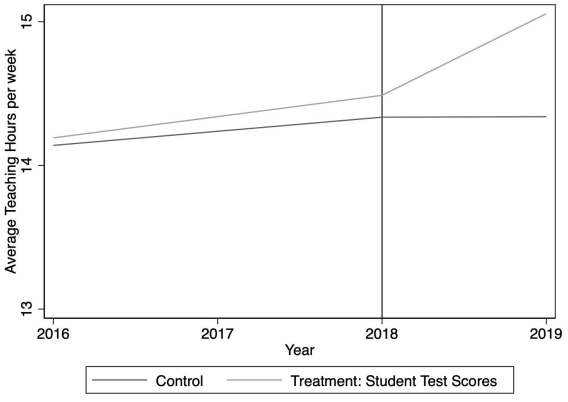
approximately only 16.8% of female teachers within the whole population, with the most female teachers in the treatment group on student test scores. Prior to the experiment, the average total hours per week is approximately 24.047 hours, which aligns with the elementary educational programs in rural areas which only lasts 4-5 hours a day (World Bank, 2008). It is also evident that there is a large variation in the workload of rural primary school teachers in Indonesia, leading to a workload of less than the minimum of 18 hours per week (World Bank, 2008).

When comparing pre and post treatment periods, it can be seen that the average attendance as well as average of both Indonesian language and mathematics test scores increase between baseline and endline observations in all groups. Table 2 also shows that groups with performance pay increase mean attendance from 77.1% to 78.5% (column 2), Indonesian language test scores increase from 39.252 to 65.212 (column 3) and mathematics test scores increase from 38.123 to 56.80 (column 4).

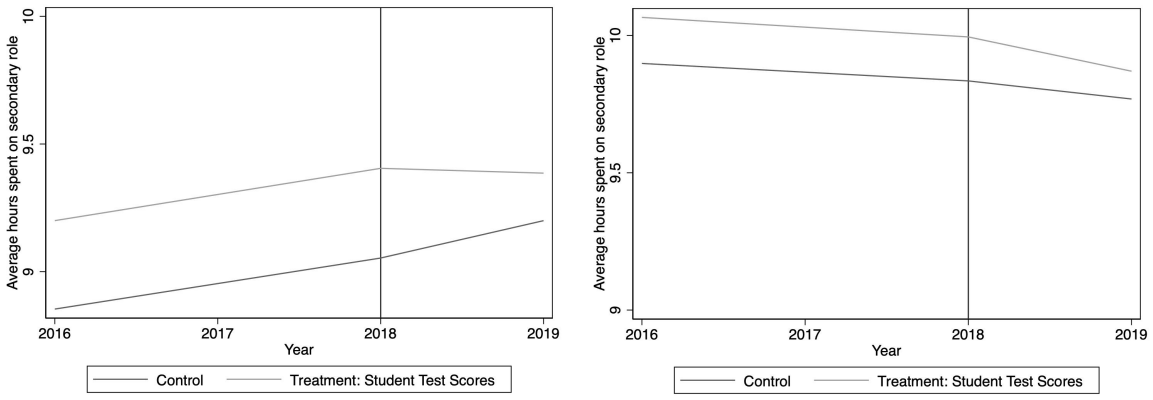
It is also interesting to see that the average teacher attendance, Indonesian language test scores and mathematics test scores also weakly increase in the control group throughout the program implementation. This potentially shows that there are possibilities of a spill over effect to the control group teachers which weakly increases the performance measures in 2019. If not, a potential bias within the control group that is caused by the implementation of the program itself. Since test scores are only taken during the end of 2018 and 2019, teachers also may put more effort during that period of time and focus on the results of the one particular test instead of a well-rounded curriculum and teaching performance. It is also important to note that the treatment group based on teachers' attendance has a small sample size compared to the other groups. This might bias the results as the groups have unequal variances.



(A)

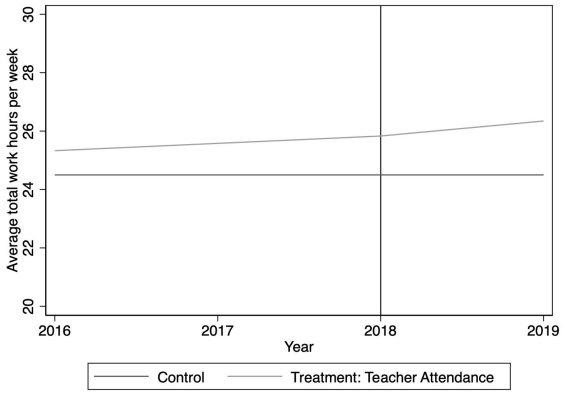


(B)

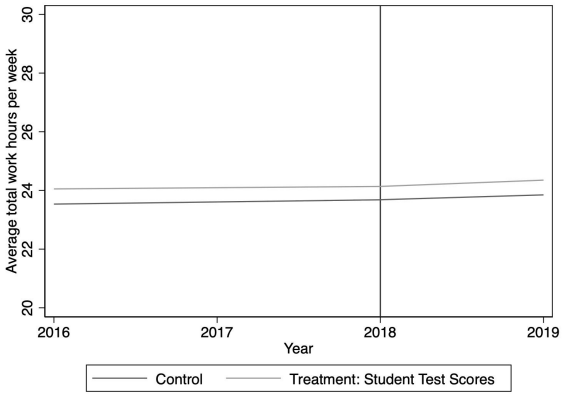


(C)

(D)



(E)



(F)

Figure 2. (A) and (B) Trends of average hours spent on teaching per week pre and post treatment period; (C) and (D) Trends of average hours spent on secondary role per week pre and post treatment period; (E) and (F) Trends of average hours working per week pre and post treatment period

Figure 2 shows the trends of the variables average hours spent on teaching per week, average hours spent on secondary roles per week and average total work hours per week through the panel data. As shown in Figure 2, with a trend between the treatment and control groups. In (A) and (B) we can see that teaching hours increase through both incentives, however, the increase is much larger for the treatment group that is incentivized through student test scores. In both (C) and (D), we can see that the average time spent on secondary roles decrease, however, again the effect seems much greater on the treatment group that is incentivized through student test scores. Specifically in (C), we can also see that the average time spent on secondary roles actually increases for the control group. We can also see that in (F), the total average work hours per week does not increase significantly for the treatment group. However in (E), when the treatment group is incentivized through teacher attendance, the total work hours increase compared to the control group. This is logical as there might be higher monitoring and teachers are then physically attending school more than previously.

Whereas with student test scores, teachers may spend the same amount of time, with more hours focused on test-related activities.

4. Main results

4.1 Effect of teachers performance pay on teacher attendance and student test scores

Table 3.

The effects of teachers performance pay on teacher attendance and student test scores

	Teacher attendance (1)	Subjects	
		Bahasa Indonesia test score (2)	Mathematics test score (3)
Treatment	.0137 (0.005)***	3.685 (0.587)***	2.544 (0.591)***
Time	-0.001 (0.004)	7.841 (0.454)***	8.594 (0.467)***
Treatment *X* Time	-0.004 (0.006)	9.747 (0.633)***	4.465 (0.637)***
Constant	0.757 (0.003)***	35.567 (0.430)***	35.580 (0.442)***
Number of observations	6,797	31,448	31,448

Note: Teacher attendance is measured in percentages of attendance per month. Both Bahasa Indonesia and mathematics test scores are measured on a scale of 0-100, as determined in the Indonesian educational system. For robustness check, an F-test concluded that all regressions were statistically significant. Robust standard errors are reported in parentheses. *, ** and *** highlights significance at the 10%, 5% and 1% levels respectively.

Table 3 shows that the treatment increases Bahasa Indonesia and mathematics test scores by 9.747 and 4.465 test points respectively. This is statistically significant at a 1% significance level. However, it does not significantly impact teacher attendance. This is aligned with Glewwe et al. (2010) showing that there is little impact of performance pay on teacher attendance in Kenya, as well as mixed results in other previous empirics (Leigh, 2012).

As part of this program, the goal from the implementation of performance pay in rural areas is not only aimed for student performance, but also to reduce absenteeism of teachers. However, it can be seen that teacher attendance has no impact, aligning with results found by Glewwe et al (2010). Using teachers' attendance as a performance measure also does not change the quality of teaching.

It could be that teachers go to classrooms but not spend enough time teaching and conducting the curriculum. There are also other social activities that teachers may engage in at school which result in the multitasking problem.

With student test scores as the performance measure, teachers would be motivated to find ways to help students in performing better at tests. Teachers are also evidently more altruistic in teaching as it provides a positive spill over to the overall human capital of their region (Palta, 2019). Teachers would then also spend more effort and time in teaching related activities, which includes teaching classes, test drills and assessments. It raises the question of whether there is a multitasking problem between teaching activities compared to secondary role activities.

4.2 Multitasking moral hazard

Table 4.

The effects of performance pay based on teachers attendance towards teacher's time spent on teaching and secondary roles

	Dependent Variable			
	Hours spent on teaching		Hours spent on secondary roles	
	Incentive performance measure: teachers attendance			
	(1)	(2)	(3)	(4)
Treatment	0.064 (0.117)	0.050 (0.117)	0 .346 (.013)***	0.386 (0.024)***
Time	0.188 (0.061)***	-0.221 (0.197)	0 .273 (0.222)***	0.722 (0.139)***
Treatment *X* Time	0.130 (0.128)	-0.245 (0.197)	-0.077 (0.043)*	-0.573 (0.198)***
Teacher age		-0.011 (0.006)*		0.010 (0.005)**
Female teacher		-0.148 (0.088)		-0.020 (0.054)
Teacher certification		0.344 (0.069)***		-0.263 (0.080)***
Intercept	15.309 (0.055)***	16.006 (0.282)***	8.853 (0.013)***	8.503 (0.212)***
Number of observations	6,797			

Note: "Teaching" (columns 1 and 2) include teaching preparation, curricular teaching, and assessments; whereas "Secondary roles" (columns 3 and 4) include extra-curricular activities, school management activities, personal development and community-building activities (such as parents meetings). For robustness check, an F-test concluded that all regressions were statistically significant. Robust standard errors are reported in parentheses. *, ** and *** highlights significance at the 10%, 5% and 1% levels respectively.

Table 4 shows that a performance pay based on teachers' attendance is more likely to reduce hours spent on secondary roles by 0.573 hours (column 4). This is statistically significant at a 1% significance level. On the other hand, it leaves teaching hours unaffected through negative insignificant results. As this treatment group is incentivised through teachers' attendance, the total hours worked does not matter to them and it only matters if they attend. Hence, it is logical that teachers tend to leave teaching activities unaffected at the cost of other secondary roles. This further reduces the total number of hours worked.

In both teaching and secondary roles, hours spent are also affected by the teacher's age and certification. Teachers' age affects hours spent on teaching negatively by 0.011 hours (column 2), whilst affecting secondary roles positively by 0.010 hours (column 4). This is aligned with Omer, Murray, & Rushton (1989) that also concluded that the older teachers are, the more likely it will reduce teachers' performance. This shows with the previous literature that with older teachers' age, it significantly reduces teaching performance (Homer, Murray, & Rushton, 1989). Teachers are then allowed to reduce time to work for teaching, and increase work in secondary roles. Certification is also more likely to increase both hours spent on teaching by 0.344 hours (column 2) and reduce hours spent on secondary roles by 0.020 (column 4). Teachers who are certificated may have higher intrinsic motivation to conduct teaching rather than their secondary roles.

Table 5.

The effects of performance pay based on student test scores towards teacher's time spent on teaching and secondary roles

	Dependent Variable			
	Hours spent on teaching		Hours spent on secondary roles	
	Incentive performance measure: student test scores			
	(1)	(2)	(3)	(4)
Treatment	0.052 (0.060)	0.045 (0.060)	0.168 (0.032)***	0.184 (0.032)***

Time	0.198 (0.049)***	0.346 (0.082)***	-0.097 (0.025)***	0.010 (0.049)
Treatment *X* Time	0.377 (0.063***	0.310 (0.094)***	-0.036 (0.034)	-0.398 (0.065)***
Teacher age		-0.005 (0.003)*		0.010 (0.002)***
Female teacher		-0.008 (0.049)		0.044 (0.028)
Teacher certification		0.114 (0.040)***		-0.132 (0.028)***
Intercept	14.140 (0.046)***	14.322 (0.127)***	9.899 (0.023)***	9.461 (0.080)***
Number of observations		31,448		

Note: "Teaching" (columns 1 and 2) include teaching preparation, curricular teaching, and assessments; whereas "Secondary roles" (columns 3 and 4) include extra-curricular activities, school management activities, personal development and community-building activities (such as parents meetings). For robustness check, an F-test concluded that all regressions were statistically significant. Robust standard errors are reported in parentheses. *, ** and *** highlights significance at the 10%, 5% and 1% levels respectively.

For the second treatment group, Table 5 shows that a performance pay based on student test scores is more likely to increase hours spent on teaching tasks by 0.310 hour (column 2) and reduce hours spent on secondary roles by 0.398 hour (column 4). Both results are statistically significant at a 1% significance level. This is aligned with the performance measure itself, as teachers are incentivized to provide higher-quality teaching or spending more time preparing for student tests. Whereas with teacher attendance, there are no performance measures related to teaching and thus, would not need to spend more time teaching. This shows evidence of multitasking and that teachers are then spending more time on activities related to teaching, at the expense of secondary roles.

Similar to the first treatment group, teachers' division of time spent in the second treatment group are also affected by the teacher's age and certification. Teachers' age reduces hours spent on teaching by 0.005 hours (column 2) and whilst affecting secondary roles positively by 0.010 hours (column 4). Certificated teachers also significantly increases time to work for teaching by 0.114 (column 2), and reduces work in secondary roles by 0.132 hours (column 4). Teachers who are certificated indicate that they have better skills, but also the motivation to conduct teaching work. Additionally, teaching activities also indicate personal utility, unlike tasks from the secondary role. However, secondary roles are often the best option for teachers in rural areas to receive higher

earnings. With the data sample also showing that the basic pay from schools for teachers with secondary roles are lower than the national average, it could be that they are more inclined to have these roles to have at least additional income apart from the regular teaching.

5. Discussion

The findings the paper found aligns with previous research around the multitasking hazard that arises from the performance-pay mechanism. This study found that through the incentive based on student test scores, teachers increase hours spent on teaching tasks and reduce hours spent on secondary roles. This aligns with the model presented by Holmstrom & Milgrom (1991). It is evident that teachers would then spend more time on teaching-related activities at the cost of secondary roles, which is organisationally very important for the school as an institution.

Empirical evidence of multitasking problems arising is seen also in a wide variety of industries including education. Multitasking problems are more likely to arise in knowledge jobs, or jobs which require tasks are ambiguous ex ante as well as generally exercise discretion and autonomy in their jobs (Bartel, 2017). This aligns with teachers as they generally have certain tasks with autonomy on how to implement it. For example, the Indonesian education system runs with a national curriculum that all public schools will have to conduct, however, teachers have the autonomy to design the class sessions, homework and method of sharing such knowledge is up to the teachers themselves.

Neal (2009) argues that an optimal performance-pay scheme should include multiple outcome measures, each of which are adjusted for the composition of the student and schools. However, there are always unverifiable performance measures, especially in the public sector like education. If the metrics are too difficult, it could also be too complicated for teachers to understand and requires additional school costs such as training and process in determining appropriate weight of metrics to ensure efficient results (Bartel, 2017). Lazear (2003) also argues that in the case of teacher performance pay, if selection effects are positive and incentive effects are zero, then teachers are unable to increase their effort level. This optimal case of performance pay would still boost student performance, but the current selection effects are believed to be not too high due to lack of teacher workforce in those regions (Shikalepo, (2020). Similar to the intervention design in this experiment, Benabou & Tirole, (2016) also saw that bonus caps can also theoretically help balance the incentives and teachers behavior, yet, it still may generate the teacher's own set of distortions. These distortions are also seen through the multitasking problem evident in this data sample. Even though the reduction of time spent on secondary roles based on the two treatment groups is not too large

compared to the gain towards student performance, it still may have potential long-run challenges. With unverifiable performance measures, the costs of multitasking could be harmful for the school itself.

When referring back to the multitasking problem, the optimal incentive scheme for the contractible dimension depends critically on the extent to which the two tasks are substitutes or complementarity (Kaarboe and Siciliani, 2011). In the case of secondary roles, the two tasks of teaching and secondary roles are therefore substitutes. With this, there arises some circumstances showing that this can be optimal. Empirically, these optimal conditions are found from Kaarboe and Siciliani (2011) where tasks are independent so that an increase in one does not come at the cost of the other. However, when referring back to the context of the data, it cannot be seen that the tasks are independent as it does come at the cost of the other in terms of losing their jobs entirely. Another condition is that if there are some elements of substitutability or if the value of the non-verifiable tasks is perceived by the teacher to be so low, that the marginal gains from an increase in the verifiable quality of teaching overcome the losses from a reduction in the non-verifiable secondary role.

In addition, focusing on the context of rural areas, the nature of the secondary jobs may be very important for the school and community itself, which also influences the holistic teacher performance. Certain performance measures may be largely influenced by certain school and village characteristics that are not taken into account. Moreover, there is evidence of the effects of performance-pay that are varied with more external characteristics or events within different firm branches (Griffith & Neely, 2009).

In general, the losses caused by performance pay might outweigh the benefits. Other alternatives of performance-inducing incentives can definitely be explored, as it is also evident that employees' attitude towards performance pay is also mixed (Leigh, 2012). Some examples of other alternatives include career and personal development, job redesigning, promotions, as well as providing other non-monetary incentives. However, such incentives are also not contractible (Holmstrom, 2016). Hence, there is still mixed evidence for which the benefits from performance-pay outweigh the losses. The most optimal design of merit based pay is thus still a large debate and more evidence in different contexts should be evaluated.

5.1 Robustness checks

The paper uses robust standard errors. To check for robustness, clustered standard errors at the treatment level are explored. In this case, we assume that treatment assignment is correlated within each group. However, it confirms that the results are similar with the ones presented through this paper and serves evidence for robustness. Throughout the model, it is also ensured to check for different specifications, including removing and adding potential control variables. It is found that results are similar in terms of significance and direction, but differs in terms of size. Hence, it can be concluded as robust and showing evidence of structural validity.

The definition of the covariates are also quite clear, with monitoring using real technology for teacher attendance. Thus, it is robust in this setting. However, it can be argued that with student scores, the characteristics of the sample would impact the results. As the data sample only focuses on primary school students, students' learning should not be evaluated only through test scores but also other skills and attitudes that are not captured through the student scores. A more holistic teacher performance measure may result in different treatment effects.

Although the data provides interesting evidence implying a multitasking moral hazard affected by the performance pay measure, the Difference-in-Difference model could not take into account time-varying factors that differ between the treatment and control group. With data on pre-treatment activities limited to one year, it is impossible to assess whether trends are similar between the groups. If there are additional periods before the intervention, it is essential to plot and test for similar trends before the intervention takes place. The data sample itself only consisted of three years with the survey only taking into account once a year, meaning that it would not fully provide evidence for the parallel trends assumption. This implies biassed results.

For further research, it would be valuable to take into account certain measures for performance of teachers in secondary roles. Some examples include how many reports are created, how many hours spent on managing extracurricular activities, how many community events are hosted and many more. Other than that, it can also be measured through school-level performance, such as financial performance, extracurricular achievements, and governmental accreditation.

6. Conclusion

To conclude, firms and organisations should take into account these potential losses as well as long-term challenges to ensure a sustainable wage contract. In an imperfect world, there is still various unverifiable information and unmeasurable variables, leading to omitted variable bias from a

performance-pay wage contract. As shown in this paper, the effectiveness of the performance pay mechanism towards teachers attendance and student test scores show mixed results. However, when testing for a potential multitasking moral hazard to arise, we saw evidence for teachers who had two roles. These two roles are vital for the outcome of the performance-pay incentive, including the performance measure itself, but also secondary roles that are vital for the management and other aspects of educational content. These potential losses should then be taken into account for further analysis. This will ensure an optimal circumstance where the organisation's desired outcomes are still achieved. Despite various limitations that could be found, the findings add evidence towards potential challenges to arise within performance-pay for public sectors and 'knowledge workers' like teachers. This will also support further experimentation on the optimal performance-pay design as well as further testing the effectiveness of such models for other sectors and regions, including further research on different variations of incentive design for firms.

Bibliography

Afkar, R., Luque, J., Nomura, S., & Marshall J. (2018). Revealing How Indonesia's Substantial Governments Spend Their Money on Education. World Bank Publications.

Al-Samarrai, S. (2014). Assessing the Role of the School Operational Grant Program (BOS) in Improving Education Outcomes in Indonesia. The World Bank.

Andrabi, T., & Brown, C. (2022). Subjective versus Objective Incentives and Employee Productivity. *Working Paper.*

Apostolova-Mihaylova, M., Cooper, W., Hoyt, G., & Marshall, E. C. (2015). Heterogeneous gender effects under loss aversion in the economics classroom: A field experiment. *Southern Economic Journal, 81*(4), 980-994.

Asian Development Bank. (2022). Teacher Absenteeism in Indonesia: Policy Brief.

Baker, G., Gibbs, M., & Holmstrom, B. (1994). The Wage Policy of a Firm. *The Quarterly Journal of Economics, 109*(4), 921–955.

Balmaseda, F. (2009). Uncertainty, Pay for Performance, and Asymmetric Information. *Source: Journal of Law, Economics, & Organization, 25*(2), 400-441.

Bandiera, O., Fischer, G., Prat, A., & Ytsma, E. (2021). Do women respond less to performance pay? Building evidence from multiple experiments. *American Economic Review, 3*(4), 435-54.

Bartel, A. (2017) Multitasking at work: Do firms get what they pay for?. *IZA World of Labor, 362.*

Benabou, R., & Tirole, J. (2016). Competitive Pay, Screening, and Multitasking. *Journal of Political Economy, 124*(2)

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50–58.

Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India (No. w24678). *National Bureau of Economic Research.*

Criado-Perez, C. (2019). *Invisible women: Data bias in a world designed for men.* Vintage.

De Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2018). Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia. *The Quarterly Journal of Economics, 133*(2), 993-1039.

Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretative Review. *The Journal of Human Resources*, *37*(4), 696–727.

Dohmen, T., & Falk, A. (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review*, *101*(2), 556-90.

Foss, N. J., & Laursen, K. (2005). Performance pay, delegation and multitasking under uncertainty and innovativeness: An empirical investigation. *Journal of Economic Behavior & Organization, 58*, 246–276.

Fryer, Jr., R. G., Levitt S. D., List, J., & Saddoff, S. (2018). Enhancing the Efficacy of Teacher Incentives through Framing: A Field Experiment. *Harvard University.*

Gaduh, A., Pradhan, M., Priebe, J., & Susanti, D. 2020. Scores, Camera, Action? Incentivizing Teachers in Remote Areas. *RISE Working Paper Series.*

Glewwe, P., Kremer, M., Moulin, S., & Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya. *Journal of Development Economics, 74(1)*, 251-268

Glewwe, P., Nauman, L., & Michael K. (2010). Teacher Incentives*. American Economic Journal: Applied Economics, 2(3), 205-27.*

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics, 118*(3), 1049-1074.

Gorman, E. H., & Kmec, J. A. (2007). We (Have to) Try Harder: Gender and Required Work Effort in Britain and the United States. *Gender and Society, 21*(6), 828–856.

Griffith, R., & Neely, A. (2009). Performance Pay and Managerial Experience in Multitask Teams: Evidence from within a Firm. *Journal of Labor Economics, 27*(1), 49-82.

Hellmann, T., & , Thiele, V. (2011). Incentives and Innovation: A Multitasking Approach. A*merican Economic Journal: Microeconomics, 3*(1), 78-128.

Hanushek, E. A., & Rivkin, S.G. (2006). Teacher Quality. *Handbook of the Economics of Education, 2,* 1052-1078.

Hibbs Jr, D. A., & Locking, H. (2000). Wage dispersion and productive efficiency: Evidence for Sweden. *Journal of Labor Economics, 18*(4), 755-782.

Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics, 10*(1), 74–91.

Harris, M., & Holmstrom, B. (1982). A Theory of Wage Dynamics. *The Review of Economic Studies, 49*(3), 315–333.

Holmstrom, B., & Milgrom, P. (1987). Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica, 55*(2), 303–328.

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

Jacobson, S. L. (1989). The effects of pay incentives on teacher absenteeism. *Journal of Human Resources,* 280-286.

World Bank. 2008. Teacher Employment and Deployment in Indonesia : Opportunities for Equity, Efficiency and Quality Improvement. *World Bank.*

Jones, D., Tonin, M., and Vlassopoulos. (2018). Paying for what kind of Performance? Performance Pay and Multitasking in Mission-Oriented Jobs. *CESifo Working Paper Series 7156.*

Kaarboe, Oddvar & Siciliani, Luigi. (2008). Multitasking, quality and pay for performance. *Working Paper in Economics, University of Bergen 07/08.*

Kitzmüller, L. (2022). *Not Merely Averages: Using Machine Learning to Estimate Heterogeneous Treatment Effects (CATE, BLP, GATES, CLAN)*. Medium.

Lavy, V. (2007). Using Performance-Based Pay to Improve the Quality of Teachers. *The Future of Children, 17*(1), 87–109.

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

Leigh, A. (2012). The Economics and Politics of Teacher Merit Pay. *Economic Studies, 59*, 1, 1–33.

Lemieux, T., MacLeod, W. B., & Parent, D. (2009). Performance pay and wage inequality. *The Quarterly Journal of Economics, 124*(1), 1-49.

Leung, M. D., California, U. of, Liang, C., Lourie, B., & Zhu, C. (2021). Heterogeneous gender effects of a firm's environmental social governance efforts on employee turnover. *Academy of Management Proceedings.*

Manning, A., & Saidi, F. (2010). Understanding the gender pay gap: What's competition got to do with it? *Industrial and Labor Relations Review, 63*(4), 681–698.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, incentives, and complementarities in education: experimental evidence from tanzania. *The Quarterly Journal of Economics, 134*(3), 1627–1673.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy, 119*(1), 39-77.

Nafilyan, V. (2020). Commuting Behaviours and the Gender Pay Gap. *King's College London - King's Business School; Government of the United Kingdom - Office for National Statistics.*

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

Oostendorp, R. H. (2009). Globalisation and the Gender Wage Gap. *The World Bank Economic Review*, *23*(1), 2009, 141–161.

Palta, Aysegul. (2019). Examining the Attitudes and the Opinions of Teachers about Altruism. *Universal Journal of Educational Research, 7,* 490-493.

Podgursky, M., & Springer, M. G. (2007). Credentials versus performance: Review of the teacher performance pay research. *Peabody Journal of Education, 82(*4), 551-573.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of policy analysis and management, 26*(4), 909-949.

Rica, S de la., Dolado, J. J., & Vegas, R. (2015). Gender Gaps in Performance Pay: New Evidence from Spain. *Annals of Economics and Statistics*, *117/118*, 41-59.

Sherry, T. B., Bauhoff, S.,, & Mohanan, M.,. (2017). Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. American Journal of Health Economics, 3(2).

Shikalepo, Elock. (2020). Improving the Quality of Education at Rural Schools in Namibia. *4.* 62-68.

Shikalepo, Elock. (2020). Challenges Facing Teaching at Rural Schools: A Review of Related Literature. *4*, 2454-6186.

Soetono, B. (2021). Implementation Completion and Results Report (ICR) Document - INDONESIA: IMPROVING TEACHER PERFORMANCE AND ACCOUNTABILITY (KIAT Guru) Phase 2 - P167216. The World Bank.

Syukri, M. (2021). Gender Equality in Indonesian New Developmental State: The Case of the New Participatory Village Governance. Jakarta: Smeru Research Institute.

Weibel, A., Rost, K., & Osterloh, M. (2010). Pay for Performance in the Public Sector—Benefits and (Hidden) Costs. *Journal of Public Administration Research and Theory, 20*(2), 387–412.

Wodon, Q., & Petroni, S. (2017). The rippling economic impacts of Child marriage. *The World Bank Blogs.* Retrieved from https://blogs.worldbank.org/education/education/rippling-economic-impacts-child-marriage

Wood, C., & Kaszubowski, Y. (2008). The Career Development Needs of Rural Elementary School Students. *The Elementary School Journal, 108*(5).

World Bank. Improving Teacher Performance and Accountability (KIAT Guru) Impact Evaluation, Baseline Survey, 2016. Ref. IDN_2016_KIAT_GURU_v01_M_id.

World Bank. Improving Teacher Performance and Accountability (KIAT Guru) Impact Evaluation, Process Monitoring, 2016-2018. Ref. IDN_2016_2018_KIAT_GURU_v01_M_id.

World Bank. Improving Teacher Performance and Accountability (KIAT Guru) Impact Evaluation, Endline Survey, 2018. Ref. IDN_2018_KIAT_GURU_v01_M_id.

Zhang, J., Jin, S., Torero, M., & Li, T. (2018). Teachers and Urban-Rural Gaps in Educational Outcomes. *American Journal of Agricultural Economics, 100*(4), 1207-1223.

Appendix

Appendix 1

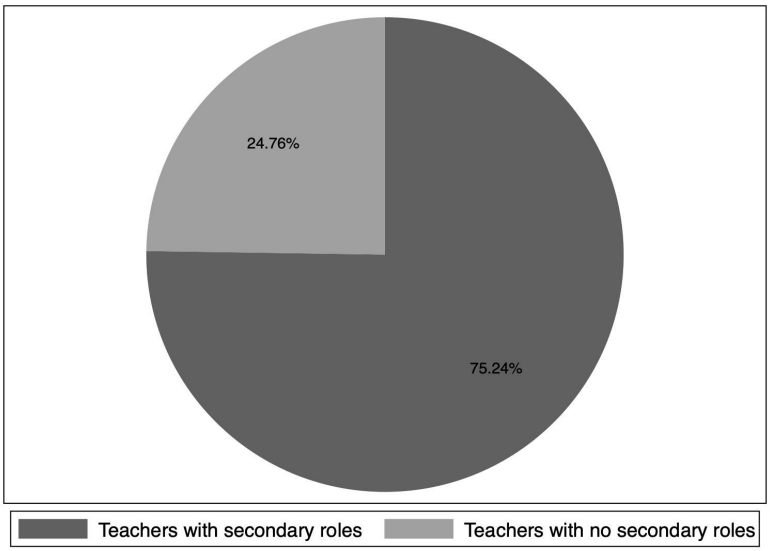


Figure 3. Ratio between teachers with secondary roles and teachers with no secondary roles in all samples.

Appendix 2

Table 2

Description and descriptive statistics of types of secondary roles of teachers

Secondary roles	Description of secondary roles
Librarian	Librarian is responsible for managing the school's library and its. The job tasks include managing collection of book inventory, cataloguing, book circulation and borrowing system, and running activities such as book clubs.
Homeroom teacher	Elementary Homeroom Teacher is responsible to supervise classrooms and provide students guidance. The job tasks include monitoring student educational progress, responsible for communication with parents, providing students advice and also observing behaviour to ensure alignment with school code.
School committee	The school committee is responsible for larger strategic decisions and ensuring efficient school operational systems. The job tasks include managing school resources, conducting community events with the school such as flag ceremony events and national holiday events, ensuring policies, hiring and monitoring school budgets.
Dapodik (data) operator	The data operator is responsible for maintaining school informatics in the national database. The job tasks include managing and reporting school informatics into the national database as well as managing the use of technology in the school.

Secondary roles	Description of secondary roles
Extracurricular instructor	Extracurricular instructors are responsible for the extracurricular activities within the school. The job tasks include designing, managing and executing the extracurricular activity assigned. Such activities may include, but not limited to, physical education, music, arts and learning about religion.
Others	Other secondary roles include, but not limited to, parent-teacher coordinator and counsellor. Scope of work may differ per individual.

Appendix 3

Table 3.

Incentives sample

	Incentives			
	None	Teacher attendance	Student test scores	Total
Number of schools	67	31	135	233
Number of teachers	471	185	766	1,422
Number of students	8,365	1,638	13,485	23,488