

Improving Teachers and School Leadership in Indonesia – Impact Evaluation of Guru Penggerak Program at the Primary Level

Indah Shafira Zata Dini, World Bank, Indonesia
Jacobus Cilliers, Georgetown University, United States
Noviandri Khairina, World Bank, Indonesia
Noah Yarrow, World Bank, United States

The Asian Conference on Education & International Development 2026
Official Conference Proceedings

Abstract

Indonesia faces a persistent shortage of qualified school principals and instructional leaders, as the country advances the Merdeka Belajar (“Emancipated Learning”) reforms. To strengthen both student-centered teaching and the leadership pipeline, the Ministry of Education launched Pendidikan Guru Penggerak (PGP), a national-scale professional development program. This study evaluates the pedagogical and leadership impacts of PGP Batch 5 using a Regression Discontinuity Design (RDD), comparing 350 primary school teachers scoring just above and below the eligibility threshold across five districts in Java. Drawing on the World Bank’s TEACH instrument and Ministry classroom observations, teacher and principal surveys, student and peer-teacher surveys, and three rounds of phone-based implementation monitoring, we find statistically significant improvements in instructional quality. Overall teaching quality increased by approximately 0.8 to 1.3 standard deviations (SD) across core instructional domains. The largest gains were observed in classroom culture and positive discipline, followed by providing a supportive learning environment, positive behavioral expectations, lesson facilitation, and promotion of critical thinking skills. PGP graduates demonstrated early signs of leadership development, they were more likely to initiate peer-learning and coaching activities. While teachers reported greater leadership confidence, principals did not yet observe measurable school-level changes, consistent with the short post-training observation window. No effects were found on socio-emotional competencies or differentiated instruction. Findings are internally valid for teachers near the cutoff in the five study districts but may not generalize nationally. Overall, the evidence suggests that structured, scalable professional development can substantially improve teaching quality and contribute to leadership development.

Keywords: teacher professional development, Guru Penggerak Program, student-centered learning, education reform, Regression Discontinuity Design

iafor

The International Academic Forum
www.iafor.org

Introduction

Understanding the Context: Indonesia's Education System

Indonesia has a large and complex education system involving both centralized and decentralized systems under multiple ministries. With over fifty-three million students enrolled in basic and secondary education and over three million teachers and education personnel, the education system is overseen by two ministries (Ministry of Education, Culture, Research, and Technology [MoECRT] & Ministry of Religious Affairs [MoRA], 2023). The Ministry of Education, Culture, Research, and Technology (MoECRT) makes and coordinates the implementation of education policies for primary, secondary, as well as higher education; and the Ministry of Religious Affairs (MoRA) manages and implements policies for the religious education sector or madrasahs across the country. MoECRT schools operate under a decentralized system implemented by provincial or local governments. Madrasahs operate in a centralized system under the MoRA following the same national curriculum as MoECRT schools, but with the ministry managing the institutions, developing religious curricula, policies, and conducting teacher training. In addition to the roles of MoECRT and MoRA, the Ministry of Home Affairs (MoHA) oversees regional governance, including teacher training at the local level and sub-national financing.

Despite significant progress in expanding access, education quality in Indonesia remains a significant challenge. For the last 30 years, Indonesia has experienced rapid progress in expanding access to primary and secondary education. However, learning quality remains a challenge. A recent World Bank report showed that only 7.8 years-equivalent of learning were acquired among youth who on average attended 12.4 years of schooling (World Bank, 2020). Results from international assessments, such as PISA 2018, show that reading literacy for Indonesian students lags behind other countries (Organisation for Economic Co-operation and Development [OECD], 2019). These findings are also aligned with results from the latest national assessment (2021), which showed that half of students do not meet Indonesia's minimum literacy standards and only a third met minimum numeracy standards (MoECRT, 2022).

Global evidence indicates that the quality of teachers and their teaching are among the most important factors in improving student learning outcomes (OECD, 2005). In 2018, the OECD conducted the Teaching and Learning International Survey (TALIS), which collected data on teachers' teaching practices and their impact on student learning outcomes across countries. The study consistently showed the impact of teacher quality on student achievement (OECD, 2020). Well-educated and skillful teachers who have deep pedagogical knowledge and are well equipped to meet diverse needs in class can deliver effective and high-quality instruction. The enhancement of teacher skills and knowledge affects the education system as teacher training continues to evolve (Barber & Mourshed, 2009). Therefore, investing in teacher education is a strategic approach to improve the overall quality of education.

The quality of teachers has been identified as one of the most significant issues in Indonesia's education sector. Data from a nationally representative survey conducted by the World Bank in 2019 found that teacher absenteeism is high, even though most absences appear to be officially excused, such as attending training or completing administrative tasks. Teacher subject knowledge was also found to be low, with only one-third of surveyed teachers able to score at least 80 percent on an assessment based on Grade 3 and Grade 4 curricula (Yarrow et al., 2020, as cited in World Bank, 2021).

There is also an urgent need to develop high-quality school principals in Indonesia, given that most school principals (72 percent) are currently aged over 50. In 2020, there were 50,135 elementary school principals aged 51–55 and 56,758 principals aged over 55 (Badan Pusat Statistik, 2020, as cited in MoECRT, 2021). This means that around one hundred thousand school principals are expected to retire within the next decade. Therefore, training for school principal candidates is a government priority. This provides an opportunity for improving school leadership at scale. Evidence shows that school leadership and management are crucial to enhancing student learning. From Brazil to India to Sweden, research reveals a significant correlation between principal management skills and student performance (World Development Report [WDR], 2018). Effective principal training focuses on improving teacher–student interaction, including providing feedback to teachers and developing action plans to enhance student performance and classroom behavior.

The COVID-19 pandemic brought an unexpected opportunity for the education sector, despite the challenges and significant impacts it caused. With limited in-person classes during school closures, the government introduced a simplified curriculum, which ultimately led to the development of the new *Kurikulum Merdeka*, aiming to provide freedom and more relevant learning experiences for students based on their needs (MoECRT, 2023). Furthermore, the pandemic accelerated the shift to online learning and the utilization of digital platforms, which in turn improved the digital literacy skills of both teachers and students. However, evidence also shows that despite these innovations, school closures caused significant learning losses, particularly among students from disadvantaged families or those with limited access to distance learning during the pandemic (World Bank, 2023).

New Wave of Reforms: Emancipated Learning

The *Merdeka Belajar* (“Emancipated Learning”) reform is one of the most significant recent policy transformations in Indonesia’s education system. Initiated in 2019, the reform emphasizes teacher autonomy, student-centered learning, and flexibility in teaching based on students’ needs (Ministry of Education, Culture, Research, and Technology [MoECRT], 2023). The reform aims to transform Indonesia’s education system into one that is more adaptive, independent, and relevant to 21st-century learning.

One major shift under the reform was the replacement of the high-stakes national examination with a new national assessment system. The government also introduced *Kurikulum Merdeka* in 2022, initially implemented through pilot schools before being expanded nationally through the *Platform Merdeka Mengajar* for teachers (MoECRT, 2023).

Student-centered learning is the core principle of the reform. This approach positions students as active participants in the learning process rather than passive recipients of information. Learner-centered education encourages problem-solving, collaboration, project-based learning, and discovery-based approaches tailored to students’ interests and capacities (Schweisfurth, 2013). Similar approaches have also been implemented successfully in countries with high-performing education systems such as Singapore and South Korea (Park, 2016).

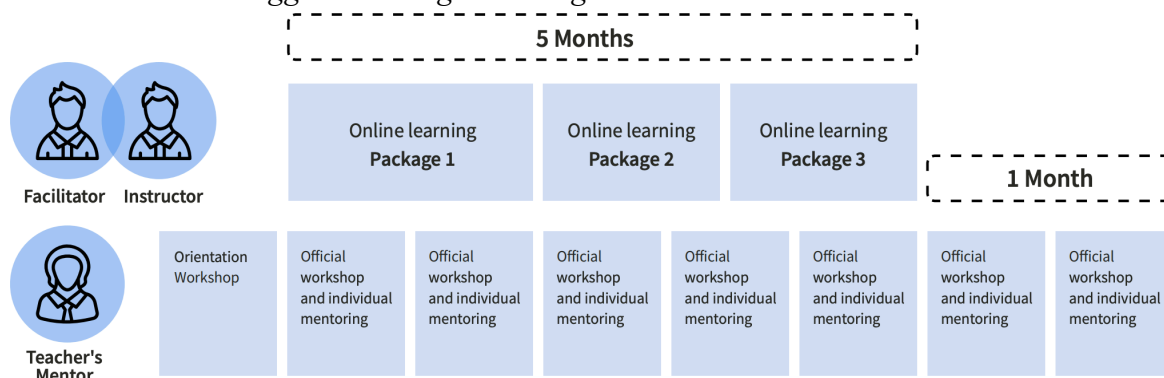
The COVID-19 pandemic accelerated the implementation of digital learning and educational innovation in Indonesia. While the pandemic caused significant learning losses, especially among disadvantaged students, it also expanded the use of online learning platforms and improved digital literacy among teachers and students (World Bank, 2023).

The World Bank is supporting the Ministry of Education, Culture, Research, and Technology (MoECRT) in its ongoing education reform effort. As part of this support, and after comprehensive discussions and careful deliberation, it was agreed that the World Bank would conduct an evaluation study specifically for the Guru Penggerak program given its importance for improving education quality in Indonesia.

Program

“Pendidikan Guru Penggerak” (PGP) program—roughly translated to instructional leaders’ program— is a teacher professional development program launched in 2020, with 2,800 teachers participating in the Batch 1 training. The program has dual objective of improving teaching practices and improve leadership by training the new cadre of school leaders who will replace the outgoing school leadership. The ministry is planning to train over 400,000 teachers by the end of 2024 and has started to expand the pace and scale of implementation. Starting this year (2023), a PGP certificate is a mandatory requirement for becoming a 1) school principal; 2) school supervisor; or 3) other leadership position in the education sector. The training is rolled out in batches. The number of teachers participating in each batch has increased from 2,800 in Batch 1 in October 2020 to 55,000 teachers who will be selected in Batch 10 by August 2023. It is estimated that about 94,000 teachers have participated in Batch 1–6 and currently ongoing batch 7, 8, and 9.

Figure 1
Pendidikan Guru Penggerak – Program Design



Source: Master Deck of Guru Penggerak program – Batch 5, MoECRT

The PGP program offers an integrated approach to enhance teachers' instructional leadership and creating student-centered learning environments. It combines online learning using a learning management system (LMS), monthly face-to-face workshops, and individual mentoring in schools over a six-month period. Online learning covers ten modules focused on student-centered learning and instructional leadership. Seven in-person workshops cover teacher competencies, community development, instructional leadership, coaching practice, collaboration for student-centered programs and sustainable self and school development. Six one-on-one in-school mentoring sessions focus on teacher competencies, evaluating classroom practice, and designing and implementing student-centered programs.

Evaluation Design

We evaluate PGP Batch 5 using a Regression Discontinuity Design, taking advantage of the fact that selection relies on an overall score given to teachers, based on their performance in a series of interviews and simulations. This estimation strategy compares the performance of

teachers who scored just below the cut-off—and who consequently did not participate in the program—with teachers just above the cutoff—who did participate in the program. We surveyed 350 current teachers and 31 former teachers who had become school principals by the time of the survey, across five districts in Java province.¹ The sample focused on teachers who were closest to the threshold for selection in their respective districts. We also observed their teaching in the classroom, and surveyed their students, another teacher in the school, and the school principal. The study also aimed to provide feedback from participating teachers and mentors, conducting phone surveys with 400 teachers and 10 mentors, to measure implementation quality.

This study reports results based on two different data collection activities (see figures below for the timeline of data collection activities). First, for the Implementation Assessment, three rounds of phone surveys were conducted during the training to capture information on the quality of implementation and teachers' perceptions of the program. Second, for the Impact Evaluation, in-person data collection took place in March and April 2023—roughly 10 months after the start of the training, and two months after the completion of the training. This include data collection using government's and TEACH classroom observation to measure the teaching practices.

Timeline of Data Collection



Descriptive Statistics

Table 1 shows basic descriptive statistics of the sample. The first column shows the mean values for those who were selected for PGP Batch 5; the second column shows the mean values for those who were not, the third column shows the mean values for all 350 sampled teachers; and the final column shows the difference between PGP and non-PGP teachers. 218 out of the 350 sampled teachers were selected to participate in the program, and 132 were not. Panel A uses administrative program data, Panel B relies on data collected during the school visits.

The average age in the sample is 38, the majority (74 percent) are female, all have a bachelor's degree but only about six percent also have a master's degree. Sampled teachers have on average 15 years of experience, and the majority are permanent civil servant teachers (62 percent).

The final column shows the difference between the PGP and non-PGP schools. As expected, those who were selected to participate received a higher score on average. They are also less likely to be female and civil servants. This might be because men and civil servant teachers received a lower score in the selection tests, on average: women received an overall score of 97, compared to men who received an overall score of 98. And civil servants received a score of 98 compared to 96 for contract teachers. If we control for the final score, then we find that women are 7 percentage points more likely to be selected, and there is no difference between civil servant and other teachers in the probability of being selected. Encouragingly, all the teachers who were selected to participate in Batch 5, according to the government records, also indicated that they did so in the survey. There is one non-selected teacher who ended up participating in Batch 7.

Table 1
Teacher Characteristics – by PGP

Variable	(1) PGP Mean/SE	(2) Control Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
<i>Administrative Data</i>				
Selection score	100.243 (0.258)	91.356 (0.318)	96.891 (0.305)	8.887***
Female	0.693 (0.031)	0.818 (0.034)	0.740 (0.023)	-0.126***
Civil servant teacher	0.683 (0.032)	0.515 (0.044)	0.620 (0.026)	0.168***
Contract teacher	0.243 (0.029)	0.341 (0.041)	0.280 (0.024)	-0.098**
Sekolah Penggerak	0.050 (0.015)	0.015 (0.011)	0.037 (0.010)	0.035*
<i>Survey data</i>				
Masters degree	0.064 (0.017)	0.045 (0.018)	0.057 (0.012)	0.019
Teacher certificate	0.830 (0.025)	0.674 (0.041)	0.771 (0.022)	0.156***
Age	37.670 (0.398)	37.765 (0.502)	37.706 (0.311)	-0.095
Years experience	15.601 (0.350)	15.530 (0.445)	15.574 (0.274)	0.071
Years of experience in school	7.583 (0.398)	8.523 (0.533)	7.937 (0.320)	-0.940
PGP—batch 5	1.000 (0.000)	0.000 (0.000)	0.623 (0.026)	N/A
PGP—batch 7	0.000 (0.000)	0.008 (0.008)	0.003 (0.003)	-0.008
Organisasi Penggerak	0.271 (0.030)	0.235 (0.037)	0.257 (0.023)	0.036
Sekolah Penggerak	0.096 (0.020)	0.076 (0.023)	0.089 (0.015)	0.021
Kurikulum Merdeka	0.321 (0.032)	0.439 (0.043)	0.366 (0.026)	-0.118**
N	218	132	350	

Notes. The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Methodology

This study focuses on evaluating the impact of the program on teacher-level outcomes, particularly teaching practices and leadership skills. Our study aims to answer the following primary research questions:

Primary Research Question

Can a government-run, in-service teacher professional development plan implemented at scale (i) improve teacher leadership and mentoring skills; and (ii) induce more student-centered teaching practices?

Will these improvements in teaching practices also spill over to other teachers and benefit the school as a whole?

In addition to these primary research questions, the study also aims to provide immediate feedback from participating teachers and mentors to further support the ministry in improving the implementation quality of PGP. Hence, a separate activity sought to answer the following secondary research questions:

Secondary Research Question

What are teachers' perceptions of the program?

What is the quality of implementation of the program, and how does this vary by district? That is, are all the elements of the program implemented correctly?

How do policies on becoming future principals influence teacher motivation to participate in the program?

This report therefore has two components: an Impact Evaluation, evaluating the impact of the program on its intended outcomes, and an Implementation Assessment, to measure implementation quality, which is to assess its accordance with the guidelines (in terms of frequency and modalities) and teacher's perception on the quality.

Empirical Approach for the Impact Evaluation—Regression Discontinuity Design

The program is evaluated using a **Regression Discontinuity Design (RDD)**, taking advantage of the fact that selection for PGP participation relies on teacher performance in interviews and simulations. This is a transparent method of converting their performance in each component into one overall score (see the section above), with only teachers who receive a high score being selected. The cut-off value for selection is different in each district since the available slots in each district are different.

The estimation strategy compares the performance of teachers who scored just below the cut-off—and who consequently did not participate in the program—with teachers just above the cutoff, who did participate in the program. Since these teachers have very similar

characteristics on average, the teachers who scored just below the cut-off are a valid counterfactual for how the PGP graduates would have performed if they were not selected.¹

The empirical strategy is slightly more. Regression analysis is used to fit a relationship between the score—what is often referred to as the running variable—and the relevant outcome indicator (e.g., teaching practices). This allows us to determine the expected outcome for a non-PGP graduate, if they were precisely at the cutoff for selection. Similarly, we use linear regression to estimate the expected outcome value for a PGP graduate at the cutoff. The difference between these two values is our estimate of the treatment effect.

When performing analysis using an RDD, there are two decisions to make. First, the size of the bandwidth: this is the number of teachers who are included in the sample for analysis. For example, one could restrict the sample to only teachers who scored one value below and one value above the cutoff; or expand it slightly to teachers who scored two values below and two values above the cutoff; or expand it even further, up to all teachers in the sample. But analysts face a trade-off here: if the sample is restricted to teachers who scored very close to the cutoff (i.e., a narrow bandwidth), then the result is the least biased. However, it is also very imprecise because of the very small sample. If the sample is expanded (for example teachers who scored between 10 below and 10 above the cutoff), then the sample is larger, and hence there is far more statistical power to detect a statistically significant impact of the program. But the larger sample introduces a risk of bias if the model is not correctly specified, because it includes teachers who scored poorly and therefore may be very different in a variety of ways from those that scored very well.

Following best practice, this report therefore shows the results for multiple different bandwidths.

A second decision is how to model the relationship between the running variable and the outcome. The easiest way is to estimate a linear relationship. This is appropriate (and the best approach) if the bandwidth is very narrow. But it becomes problematic with a wide bandwidth. In this case, it is appropriate to allow for a non-linear relationship between the running variable and the outcome.

In this report the results are therefore reported with the following specifications:

Very narrow bandwidth: The sample is restricted to teachers who scored just two points below and two points above the cutoff. This model is estimated with a linear specification for the running variable.

Narrow bandwidth. The sample is restricted to teachers who scored just three points below and three points above the cutoff. This model is also estimated with a linear specification for the running variable.

Full sample, including all surveyed teachers. This model is estimated with a quadratic specification of the running variable.

We only consider a result to be robust if the magnitude is consistently and meaningfully large under the three different estimation choices, and statistically significant for at least one of the strategies. Note that it is possible that the estimated coefficients are similar in magnitude, but only the estimation using the full sample is statistically significant, given the increased statistical power from using the full sample. We still consider this to be a robust result.

¹ The counterfactual is beneficiaries' outcomes if they did not participate in the program. It is critical have a valid measure of the counter-factual in order to estimate the causal impact.

Sample Selection – Examining the Cutoff for Selection

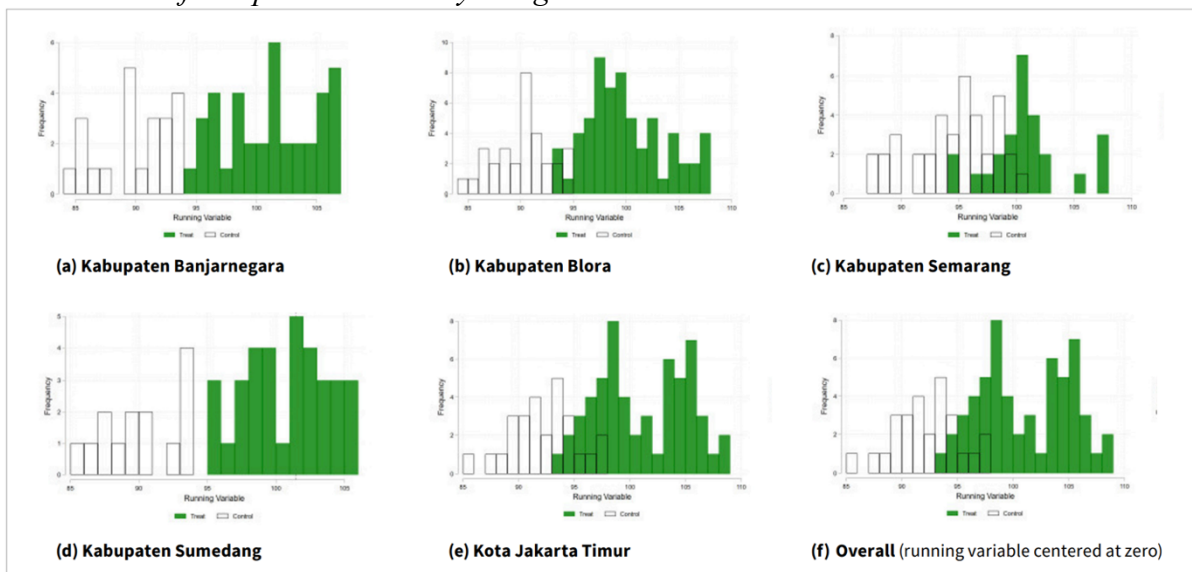
The sample is drawn from all primary-level teachers who were short-listed to participate in Batch 5. Out of these teachers, 350 teachers who were closest to the threshold for selection in their respective districts/cities, were sampled for data collection. The sampling process involved identifying five locations that had the most teachers close to the threshold, and then sampling 350 teachers whose score was closest to the threshold. The final sample had 262 treated and 138 control teachers. The sampled locations are: Kabupaten Banjarnegara, Kabupaten Blora, Kabupaten Semarang, Kabupaten Sumedang, and Kota Jakarta Timur, all are located in Java island.

Figure 2 below shows the distribution of the selection score (which we refer to as the “running variable”), broken down by whether the teacher in our sample was selected to participate in PGP or not. The treated teachers (i.e., those selected into the program) are shown in green bars, the control teachers (who were not selected into the program) are shown the clear bars. Panels (a) to (e) show the distribution separately for each district in our sample. There is some non-compliance in some of the districts: i.e., there are some teachers who scored above the threshold yet were not selected, and some who scored below the threshold and who were selected. This has some implications for the empirical strategy and also reduces the statistical power. But it does not undermine the internal validity of the results: we are still able to estimate an unbiased causal estimate of the impact of participating in the program.

Figure 2(f) shows the distribution for all the teachers, after we have “centered” the running variable. We do this in two steps. First, we determine the cut-off value for selection into the program separately for each district and teacher contract type. Second, we subtract a teacher's score by this cutoff value.² A value of zero means that they received the minimum score required for a teacher of their contract type and in their district to be selected. Figure 5(f) clearly shows how there are far more treated teachers to the right of the cut-off, compared to the left of the cut-off, which supports our evaluation design and empirical strategy.

² For example, it is 94 for all teachers in Banjarnegara and 95 for all teachers in Sumedang. In two districts the cutoff value that maximizes the power of the first stage is slightly lower for PNS teachers.

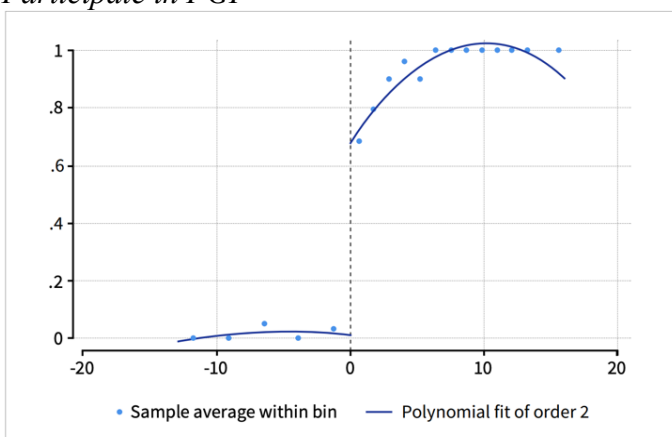
Figure 2
Distribution of Sampled Teachers by Assignment to Treatment



Note. The running variable is the teachers’ overall score in the second round of selection. In Figure (f) it is centered to be equal to zero at the cutoff for selection for a given teacher contract and district.

Figure 3 below confirms that there is a discontinuous jump in the probability of being selected at the cutoff. The proportion of teachers who were selected jumps from just above zero below the cutoff, to about 70 percent above the cutoff. This is encouraging, since we have more statistical power if there is a larger jump in the probability of being treated at the cutoff value.

Figure 3
The Relationship Between the Running Variable and the Probability of Being Selected to Participate in PGP



Note. Y axis is the proportion of teachers who were selected to participate in Batch 5. X-axis is the running variable, centered at zero. Each dot represents the mean value for teachers with a given value of the running variable.

Executive Summary of Impact Evaluation Results

First, the program was well-implemented and well-received by the teachers. Teachers rated all components of the program highly. Moreover, overall satisfaction with this program is higher than for teachers who participated in other training programs. Teachers are also more

likely to believe that this training program will greatly improve their teaching skills, with 86 percent for this program versus 53 percent for others.

One area for potential improvement is the support provided to teacher mentors. Overall, the perception of the quality of the mentors is lower than the perceived quality of the instructors or facilitators. There is also substantial regional variation, with an average score of 8.3 out of 10 in one district. Mentor teachers also provided a low rating for the quality of training that they received. Their training is exclusively online, and there is no guidance on how to use the classroom observation instrument that they are required to administer during coaching sessions.

Second, teaching practices improved substantially, especially in classroom culture and positive discipline. Overall teaching practices, measured by the Teach classroom observations instrument, increased by 0.8 to 1.1 standard deviations (SD), which suggests the teaching practices in the classrooms improved on average, and the extent of improvement is substantial. The largest improvements were in providing a supportive learning environment, followed by positive behavioral expectations, lesson facilitation, and promotion of critical thinking skills. There is also consistent evidence, from both the student surveys and MoECRT classroom observation instrument, of improvements in positive discipline. Students were less likely to report that they were scolded or punished by a teacher, less afraid to ask questions in the class, and more likely to have a class agreement or social contract. Teachers are 34 to 50 percentage points more likely to apply the “restitution process” when disciplining children, which entails reminding children of the class agreement when they misbehave, and asking children to self-reflect on what they did wrong.

However, there is no evidence of improvements in differentiated instruction, or in teachers’ socio-emotional skills in the classroom. Along with positive discipline, these were three elements of teaching practices that the program aimed to improve. This result suggests that it may be harder to change teachers’ socio-emotional skills than to shift specific instructional practices, such as applying the restitution process.

Third, there is evidence that the PGP teachers provided instructional support to other teachers in the school, but only if we restrict our sample to schools where there are no other PGP graduates. A requirement of the program is that PGP teachers establish communities of learning and provide instructional support to other teachers in the school. Schools are now more likely to have communities of learning as a result of the program, but this might be because almost 90 percent of control schools already have a community of learning. Almost 40 percent of teachers in our control group teach in schools where there is at least one PGP graduate, so the control teachers might have already benefited from the program. When restricting the sample to schools where there are no other PGP teachers, we find that the peer teachers are more likely to indicate that the PGP teacher-initiated discussions about student-centered learning, helped them map out students’ needs, asked them to self-reflect on their learning process, and helped them conduct a strength mapping exercise.

Fourth, the program improved teachers’ enjoyment of teaching, their aspirations to become school principals, and their own self-confidence as a leader. This is very encouraging, since it suggests that the new teaching practices reduced the anxiety and stress that teachers face in the classroom and increases the likelihood that these improved teaching practices will be sustained over time. However, at the time of the data collection, we don’t find evidence of

improvement in teacher's leadership abilities as measured by perception of the school principals.

The complete tables and results of this report can be accessed through official World Bank website:

<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099042224202520319>

Limitation

The positive impacts found in this study has the potential lack of generalizability due to limitation that our sample were located in districts within Java Island, which is home to most well-developed provinces and districts in Indonesia. Our sampling strategy for Regression Discontinuity Design the (RDD) meant that we only sampled teachers close to the cutoff, and it happened that the only districts that had a large enough sample of short-listed teachers who scored close to the cutoff were all in Java. Consequently, while the internal validity of our evaluation method is strong, the findings of this study cannot be generalized to all provinces and districts across Indonesia. Furthermore, the focus of this study is on CGP at the elementary school level, so that different outcomes may be found at lower or higher educational levels.

Policy Conclusions

We conclude that PGP should continue, given the high quality of implementation and its positive impact on teaching practices and leadership skills. We do, however, expect significant implementation challenges as the ministry scales up the rollout of training over the next 18 months. Our results also indicate areas for improvement, both in the quality of implementation and design of the program.

Additional Support to Mentors

While MoECRT has developed a comprehensive handbook for mentors, there is still room for improvement. In particular, additional guidance is required for the classroom observations, such as how to rate teacher competencies according to the government instrument, and what kind of feedback to provide to teachers based on the observations. This can also be achieved by enhancing mentor training to include additional pedagogical strategies and effective mentoring techniques. By equipping mentors with a broader range of knowledge and skills, they will be better prepared to support teachers in their professional growth. Lastly, mentors also need to have robust support system and access to resources to overcome challenges, share experience, and enhance their effectiveness in guiding the teachers.

Supporting Teachers in Differentiated Learning and Socio-Emotional Skills

This support can take the form of materials (guides, tools, practical tips, other resources) related to differentiated learning in the classroom or easily accessible self-paced training for teachers, while also ensuring a learning community at the school level. Support can also encompass additional mentoring sessions during the program and follow-up mentoring sessions after the program ends.

Standardize Best Practice and Provide Clear Guidelines and Procedures for Implementation of District-Level Activities

As the program scales up, it is relying more on district-level implementation, which could be variable. Clear guidelines and related resources, as well as coordination between the Ministry of Education, Culture, Research, and Technology (MoECRT) and related stakeholders, including local government and Balai Guru Penggerak, is vital for effective program execution.

Reconsider the Broad and Ambitious Objectives of the Program

The program might be overly ambitious in the wide range of teacher attitudes and skills that it hopes to change. Socio-emotional competencies, in particular, are very hard to change in adults. Moreover, the official standards for reaching competence in the different domains, as defined in the program's classroom observations instrument, are highly ambitious. A different strategy is to prioritize a smaller set of skills that are key to improving student outcomes, and setting standards that are realistically achievable by all teachers. This will be increasingly important as the program continues to scale up and is implemented in remote and under-resourced locations where the baseline teacher capacity might be weak.

Conduct Additional Research

The results of this study are not generalizable to Indonesia as a whole. Future research should:

- Expand data collection to more districts, including remote and rural locations.
- Collect data on teachers who participated in both earlier and later batches of training.
- Measure students' cognitive and non-cognitive skills.
- Conduct a prospective randomized evaluation in one of the upcoming batches, using a randomized phased in approach. This will strengthen the confidence that the program had a causal impact, increase the generalizability of the findings, and also improve statistical power.

Acknowledgements

This report was prepared by a World Bank team led by Noviandri Khairina, and comprising (in alphabetical order) Jacobus Cilliers, Indah Dini, and Noah Yarrow. Alexander Michael Tjahjadi and Ross Babineau provided support on data cleaning and analysis. Sylvia Njotomihardjo provided administrative assistance to the task team over the course of the preparation of this report. Additionally, special acknowledgment goes to Nuriza Saputra for the design of this report.

This report was developed under guidance provided by Cristian Aedo (Education Practice Manager, East Asia and Pacific Region), Achim Schmillen (Human Development Practice Leader, East Asia and Pacific Region), Shinsaku Nomura (Education Cluster Lead, East Asia Pacific Region) and Satu Kahkonen (Country Director for Indonesia and Timor-Leste).

The team is indebted to Rythia Afkar, Ezequiel Molina, Koen Geven, Lars Sondergaard and other reviewers for their valuable feedback for preparing and improving this report.

The World Bank Task Team would like to thank the Ministry of Education, Culture, Research, and Technology (MoECRT), for all the support and input provided during this research. Particularly to Prof. Dr. Nunuk Suryani, M.Pd. (Director General of Teachers and Education Personnel (GTK), Mr. Iwan Syahril, Ph.D. (Director General of Early Childhood and Basic Education), Dr. Kasiman (Acting Director of KSPSTK), Dr. Praptono (Secretary of the Directorate General of Early Childhood and Basic Education), the Setditjen GTK Team (Mila Novita, Reza Maulana Hamzah), the Pokja Team at the Directorate of KSPTK (Rita Dewi Suspalupi, Otong Kusnadi, Yudi Herman, and Agus Rohmani), as well as the Technical Team of Teachers and Education Personnel (Tim Teknis GTK).

References

- Badan Pusat Statistik. (2020). *Education Statistics Indonesia 2020*. BPS RI.
- Barber, M., & Mourshed, M. (2009). *How the world's best-performing school systems come out on top*. McKinsey & Company.
- Ministry of Education, Culture, Research, and Technology. (2021). *School principal development and retirement projections in Indonesia*. MoECRT.
- Ministry of Education, Culture, Research, and Technology. (2022). *Minister presentation on the national assessment (ANBK) result*.
- Ministry of Education, Culture, Research, and Technology. (2023). *Kurikulum Merdeka policy framework*. MoECRT.
- Ministry of Education, Culture, Research, and Technology, & Ministry of Religious Affairs. (2023). *Education statistics and EMIS 2022/2023*.
- Organisation for Economic Co-operation and Development. (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. OECD Publishing.
- Organisation for Economic Co-operation and Development. (2019). *PISA 2018 results*. OECD Publishing.
- Organisation for Economic Co-operation and Development. (2020). *TALIS 2018 results (Volume II): Teachers and school leaders as valued professionals*. OECD Publishing.
- Park, H. (2016). *Education reform and student-centered learning in South Korea*. Korean Educational Development Institute.
- Schweisfurth, M. (2013). *Learner-centred education in international perspective: Whose pedagogy for whose development?* Routledge.
- World Bank. (2020). *Human Capital Index 2020 update: Human capital in the time of COVID-19*. World Bank.
- World Bank. (2021). *Improving teaching quality and learning outcomes in Indonesia*. World Bank.
- World Bank. (2023). *Indonesia education sector assessment: Learning recovery and transformation*. World Bank.
- World Development Report. (2018). *Learning to realize education's promise*. World Bank.
- Yarrow, N., Masood, E., & Afkar, R. (2020). *Estimating teacher absenteeism and quality in Indonesia*. World Bank.

Contact emails: nkhairina@worldbank.org
ejc93@georgetown.edu
idini@worldbank.org
indahshafira29@gmail.com