



Improving Teachers and School Leadership in Indonesia

Impact Evaluation of Guru Penggerak Program at the Primary Level

Noviandri N. Khairina, Noah Yarrow,
Jacobus Cilliers, and Indah S.Z. Dini.

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

This report was prepared by Noviandri N. Khairina, Noah Yarrow, Jacobus Cilliers, and Indah S.Z. Dini.

Acknowledgement

This work is a product of the staff of The World Bank. Financial support for this work was generously provided by the Australian Government through the Australia-World Bank Indonesia Partnership (ABIP). This research product also benefited from additional financing that comes from the China-World Bank Trust Fund. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Rights and Permissions

© 2023 The World Bank
1818 H Street NW, Washington DC 20433
Telephone: 202-473-1000; Internet: www.worldbank.org
Some rights reserved

The material in this work is subject to copyright. Because The World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for non-commercial purposes as long as full attribution to this work is given.

All queries on rights and licenses, including subsidiary rights, should be addressed to World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: pubrights@worldbank.org.

Attribution

Please cite the work as follows: Khairina, Noviandri; Yarrow, Noah; Cilliers, Jacobus; Dini, Indah. 2023. Improving Teachers and School Leadership in Indonesia: Impact Evaluation of Guru Penggerak Program at the Primary Level. World Bank, Jakarta. © World Bank.

Contact Information

The authors can be contacted at nkhairina@worldbank.org, ejc93@georgetown.edu, idini@worldbank.org.

Photo Credit:

Cover and Chapter Opening: Freepik, Midjourney

Table of Contents

Table of Contents	4
Tables and Figures	4
Abbreviations and acronyms	5
Acknowledgement	5
Executive Summary	6

1. Understanding the Context: Indonesia's Education System	10
2. New Wave of Reforms: Emancipated Learning	12
3. The Program – Pendidikan Guru Penggerak (PGP)	14
Program Design, Delivery, and Curriculum	14
Selection process	16
Theory of Change	17
4. Methodology	18
Empirical Approach for the Impact Evaluation—Regression Discontinuity Design	18
Sample Selection - Examining the Cutoff for Selection	19
Limitations	21
5. Data and Instruments	22
Implementation Assessment (phone survey)	22
School Visit for Impact Evaluation	23
Instruments	23
Construction of Indicators and Indices	28
Timing of Data Collection	28
Descriptive Statistics	28
6. Results	30
Phone Survey for Implementation Assessment Results	30
Impact Evaluation Results	33
Quality of Implementation and Other Training Programs	34
Self-reflecting and Self-improving Teachers	36
Student-centered Teaching	37
Leadership: Communities of Learning, Coaching, and School vision	42
Additional Findings from This Evaluation	44
7. Conclusions and Recommendations	48

References	52
Appendix 1 – Pendidikan Guru Penggerak	53
Appendix 2 – Teach Classroom Observation Instrument:	60
Appendix 3 – Empirical Strategy	64
Appendix 4 – Additional Tables and Results	65

List of Tables

Table 1	Comparison of 'Penggerak' programs in the Merdeka Belajar reform	12
Table 2	Sample for Phone Survey and School Visit Data Collection	22
Table 3	Summary of Phone Survey Data Collection	23
Table 4	Mapping between MoECRT's groupings of final outcomes and our evaluation instruments	24
Table 5	Description of the Teach classroom observation instrument	26
Table 6	Summary of the MoECRT classroom observation instrument	27
Table 7	Teacher Characteristics - by PGP	29
Table 8	Summary of main results	33
Table 9	Quality of Implementation by District	34
Table 10	Exposure to training (by selection into Batch 5)	35
Table 11	TOC I. Self-reflection and self-development	36
Table 12	Leadership confidence and aspirations	46
Table 13	Leadership confidence and aspirations (by gender)	47
Table 14	School Principal's perception of the teacher's leadership ability	47
Table 15	TEACH Results	62
Table 16	Overview of TEACH Instrument	63

List of Figures

Figure 1	Pendidikan Guru Penggerak – Program Design	15
Figure 2	PGP Actors roles and relationships	16
Figure 3	PGP Theory of Change: Final Outcome	17
Figure 4	Location of Sample Districts	20
Figure 5	Distribution of sampled teachers by assignment to treatment	20
Figure 6	The relationship between the running variable and the probability of being selected to participate in PGP	21
Figure 7	Timeline of Data Collection	28
Figure 8	Quality of implementation for different components of the program	30
Figure 9	Distribution of perceived quality of the facilitators, instructors, and mentors	31
Figure 10	District-level variation in perceived quality of the facilitator, instructor and mentor	31
Figure 11	Mentor Training Effectiveness	32
Figure 12	How workshop can be improved	32
Figure 13	Proportion of teachers that said logistics can be improved	32
Figure 14	Teachers' five-year career plans	33
Figure 15	reflection and self-development	36
Figure 16	TOC II.1 Student-centered teaching – positive discipline	38
Figure 17	TOC II.2 Student-centered teaching – differentiated learning	39
Figure 18	TOC II.3 Student-centered teaching – socioemotional competencies	40
Figure 19	TEACH Scores	41
Figure 20	TOC III.1 Create communities of learning (restricted sample)	42
Figure 21	TOC III.2 Coaching (Restricted sample)	43
Figure 22	TOC IV.2 Vision of student-centered school development (restricted sample)	44
Figure 23	Enjoyment of teaching	45
Figure 24	TEACH Result - Time on learning	61
Figure 25	Teach Primary Score of All Areas	61
Figure 26	Distribution of Average Teach Primary Scores by Area and Element	63

Abbreviations and Acronyms

BGP/BBGP	—	MoECRT's Implementation Unit on Teacher and Education Personnel Development
CGP	—	Guru Penggerak Training Participants
EMIS	—	Education Management Information System
GTK	—	Directorate General of Teachers and Education Personnel
KM	—	Merdeka Curriculum
KSPSTK	—	School Principals, School Inspectors, and Education Personnel
LMS	—	Learning Management System
MoECRT	—	Ministry of Education, Culture, Research, and Technology
MoRA	—	Ministry of Religious Affairs
OECD	—	Organisation for Economic Co-operation and Development
PGP	—	Guru Penggerak Program
PISA	—	Programme for International Student Assessment
PNS	—	Civil Servant
PSPK	—	Center for Education and Policy Studies
RDD	—	Regression Discontinuity Design
SD	—	Standard Deviations
TALIS	—	Teaching and Learning International Survey
ToC	—	Theory of Changes
WDR	—	World Development Report

ACKNOWLEDGEMENT

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 KSPSTK (Rita Dewi Suspalupi, Otong Kusnadi, Yudi Herman, and Agus Rohmani), as well as the Technical Team of Teachers and Education Personnel (Tim Teknis GTK).

EXECUTIVE SUMMARY

Background

The most significant recent education reform in Indonesia is 'Merdeka Belajar', officially translated as “**emancipated learning**,” which aims to transform the education system into having a more student-centered approach and promote lifelong learning. The reform places a great emphasis on teacher competence, independence, freedom, and ability to teach based on student's needs, and aims to delegate more autonomy to teachers and schools for enhanced student-centered learning. The program includes many components, one of which is teacher training.

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.

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.

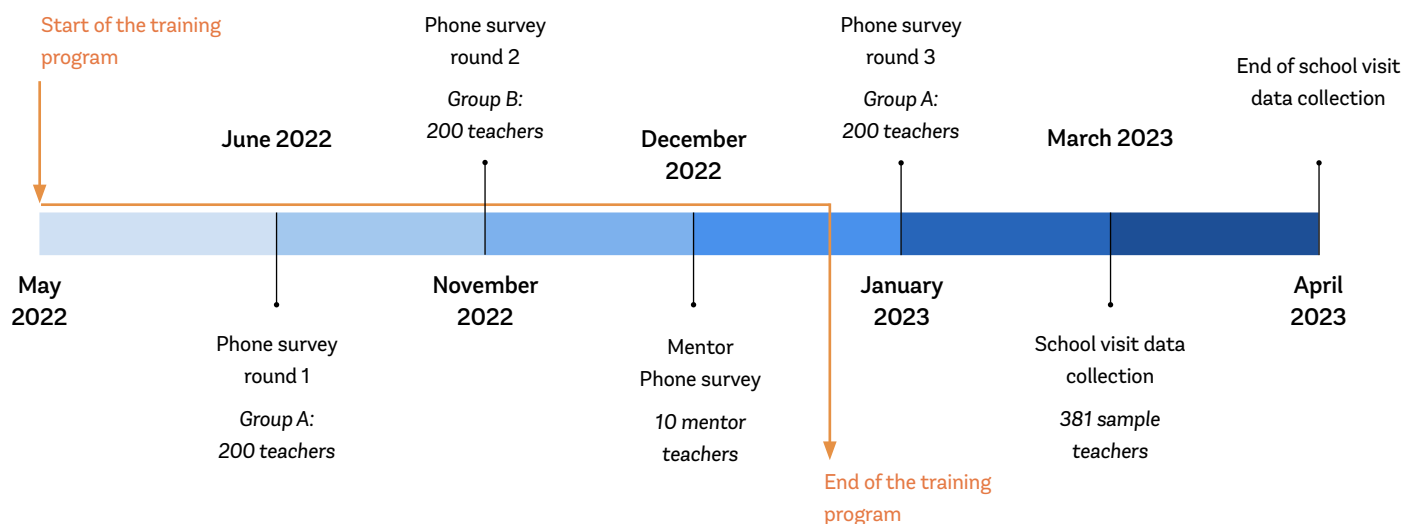
¹ Kabupaten Banjarnegara, Kabupaten Blora, Kabupaten Semarang, Kabupaten Sumedang, and Kota Jakarta Timur

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



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.

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 the regression discontinuity design (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

- 1. 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.

- 2. 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.
- 3. 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.
- 4. 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.
- 5. Conduct additional research.** The results of this study are not generalizable to Indonesia as a whole. Future research should:

A. Expand data collection to more districts, including remote and rural locations.

B. Collect data on teachers who participated in both earlier and later batches of training.

C. Measure students' cognitive and non-cognitive skills.

D. 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.

1.



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 students² enrolled in basic and secondary education and over three million teachers and education personnel, the education system overseen by two ministries. 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 report³ showed that only 7.8 years-equivalent of learning were acquired among youth who on average attended 12.4 years of schooling. Results from international assessments, such as PISA 2018, show that reading literacy for Indonesia students lags other countries.⁴ These findings are also aligned with results from the latest national assessment (2021), which showed that half students do not meet Indonesia's minimum literacy standards and only a third met minimum numeracy standards.⁵

² MoECRT Education Statistics, MoRA's EMIS 2022/2023

³ Human Capital Index, 2020

⁴ Organisation for Economic Co-operation and Development (OECD), 2018

⁵ MoECRT's minister presentation on the national assessment (ANBK) result, April 2022

Global evidence indicates that the quality of teachers and their teaching are the most important factors in improving student learning outcomes.⁶ In 2018, the OECD conducted a Teaching and Learning International Survey (TALIS) which collected data on teachers' teaching practices and the impact on student learning outcomes across countries. The study consistently showed the impact of teacher quality on student achievement.⁷ 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 and 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 World Bank in 2019 found that teacher absenteeism is high, even though most of this absence appears to be excused and officially excused (e.g. teacher was absence to attend training or completing other administrative task). Teacher subject knowledge was also found to be low, with only a third teachers surveyed able to score at least 80 per cent in an assessment based on material from the Grade 3 and Grade 4 curricula.⁸

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 are 50,135 elementary school principal age between 51-55 and 56,758 principals age over fifty-five.⁹ This means that a hundred thousand school principals are expected to retire within the next decade.¹⁰ 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 (WDR, 2018). Effective principal training focuses on how to improve teacher-student interaction, such as 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 impact it caused. With limited in-person classes during the period of school closures, the government introduced a simplified curriculum, which ultimately led to the development of the new *Kurikulum Merdeka*, which aims to provide freedom and a more relevant learning experiences for students based on their needs.¹¹ Furthermore, the pandemic engendered a shift to online learning and the utilization of digital platforms which in turn increased the digital literacy skills of both teachers and students. However, evidence shows that in spite of these innovations, the pandemic closures caused significant learning loss, particularly for students from the most disadvantaged families, or those who had limited access to distance learning during school closures.¹²

¹¹ The enhancement of teacher skills and knowledge affects the education system as teacher training continues to evolve (Barber and Mourshed, 2009). Therefore, investing in teacher education is a strategic approach to improve the overall quality of education. ¹¹

⁶ OECD, 2005

⁷ OECD, 2020

⁸ Yarrow et. al., 2020⁹ Badan Pusat Statistics, 2020 as cited in Katadata, 2021

¹⁰ Mandatory retirement age for school principals is 65 years old (Circulate Letter from Head of Civil Servant Agency No. K26-30/V.119-2/99 on October 3rd, 2017.

¹¹ MoECRT, 2023

¹² World Bank, 2023

2.



NEW WAVE OF REFORMS: EMANCIPATED LEARNING

‘Merdeka Belajar’ reform, officially translated as “emancipated learning”, is the most recent policy reform in Indonesia’s education system. The reform, initiated in 2019, emphasizes teacher competence, independence, freedom, and ability to teach based on student needs, and aims to delegate more autonomy to teachers and schools for enhanced student-centered learning. The overarching objective of this reform is to transform Indonesia’s education system for the 21st century, to be more student-centered, and to promote lifelong learning. By April 2023, the ministry had launched 24 Merdeka Belajar programs for regional education offices, schools, teachers, and students, and covering all levels of education from early childhood, primary, secondary, vocational, and up to higher education.

The education sector reform involves multiple programs targeting a variety of stakeholders in the sector. The Ministry started the reform by abolishing the high-stake national exam in 2020 and introducing new national assessments. This was followed by a series of interventions, including the ‘Penggerak’ programs: *Organisasi Penggerak*, *Guru Penggerak*, and *Sekolah Penggerak*. The three are similar

in terms of the overall objective to develop students with *Pancasila profiles*,¹³ but are different in terms of means of the intervention and duration of the program (as summarized in table below). In February 2022, MoECRT introduced a new curriculum, ‘Kurikulum Merdeka’, which had been implemented as a pilot in 2,500 schools participating in ‘Program Sekolah Penggerak’. Shortly afterwards, the Ministry enabled implementation of the new curriculum for academic year 2022/2023 by providing access to an online platform, ‘Platform Merdeka Mengajar’, for all teachers to learn about the curriculum. These are important factors in our analysis, since even teachers in the control group for this study might have benefited from other programs and interventions.

Student-centered learning, the key element of the reform, places students at the center of learning process as opposed to the traditional teacher-centered model. The concept of student-centered learning is not a recent trend in global education, as it has been discussed in the international literature for over two decades. In essence, learner-centered education refers to a pedagogical approach

¹³ The “Profil Pelajar Pancasila” refers to the profile of students who embody the values and principles of Pancasila, which is the philosophical foundation of the Indonesian state. It encompasses the characteristics, attitudes, and behaviors that reflect the ideals of nationalism, humanism, social justice, democracy, and belief in one supreme God. The “profil pelajar pancasila” aims to develop students who are morally upright, responsible, respectful, and actively contribute to the betterment of society, guided by the principles of Pancasila.

that gives learners (and demands from them) a relatively high level of ownership over the content and process of learning in accordance with their needs, capacities and interests (Schweisfurth, 2013).¹⁴ This places the learner rather than the teacher, content or system, at the heart of the teaching and learning processes.¹⁵ Learner-centered education involves student-activating approaches, problem-based learning, minimal guidance approach, discovery learning, open-ended learning environments, collaborative/cooperative learning, project-based learning, and case-based learning (Baeten et al. 2010¹⁶).

Similar approaches have been implemented in other countries with high-performing education sectors such

as Singapore and South Korea. South Korea has promoted student-centered learning since the 1990s, initiated by *The 31 May Education Reform Report* in 1995¹⁷ and later through a series of revisions to the curriculum over the following decades. An example of a student-centered initiative is the 'Creative Experimental Activity' introduced in 2009 for middle and high school students where they participate in interesting and experiential activities for 3-4 hours a week (Park 2016¹⁸). Meanwhile in Singapore, the shift toward a more student-centric learning began with the "Student-centered, Values-driven" vision articulated in 2011¹⁹, which emphasizes holistic education centered on values, social and emotional well-being and character development for every child.

Table 1. Comparison of 'Penggerak' programs in the Merdeka Belajar reform

	Program Sekolah Penggerak	Program Pendidikan Guru Penggerak	Program Organisasi Guru Penggerak
Objective	To improve student learning outcomes by improving competencies of existing school leaders and teachers.	To produce instructional leaders who will encourage student-centered learning and act as agents of transformation in the education system	Massive community empowerment effort to improve teacher and principal quality through various training models and design
Actors	School Principals, Facilitators of Sekolah Penggerak	Teachers, Mentor, instructors and Facilitators of Guru Penggerak	NGOs or development partners
Program Duration	3 years	9 months (Batch 1-4) 6 months (Batch 5-13/ present)	2 years
Target Beneficiaries	School principals, teachers, supervisors	Teachers	School principals, teachers, supervisors

Source: MoECRT presentation on Merdeka Belajar, episode 4, 5, and 7 - (<https://merdekabelajar.kemdikbud.go.id/>)

¹⁴ Schweisfurth, 2013

¹⁵ Yang and Tan, 2019

¹⁶ Baeten et.al., 2010

¹⁷ Kim, 2004

¹⁸ Park, 2016

3.



THE PROGRAM: PENDIDIKAN GURU PENGGERAK (PGP)

As part of the ongoing program of reform, MoECRT launched a new in-service teacher training program in 2020.²⁰ The program, called 'Pendidikan Guru Penggerak', is a professional development program offering training and mentoring focused on instructional leadership. It aims to empower teachers to facilitate student growth, inspire their colleagues to implement student-centered learning, and serve as role models and agents of transformation in the education system to realize the profile Pelajar Pancasila. The PGP initiative was introduced as a set of other programs: "transformational organization program" and "transformational school program" (*Program Organisasi Penggerak* and *Program Sekolah Penggerak*). The ministry is planning to train 407,400 teachers by the end of 2024. As of June 2023, the ministry has just completed recruiting 62,681 teachers to participate in PGP.²¹

A new policy and regulation provides a pathway for teachers to become principals or supervisors. The government has implemented a new regulation, Permendikbud No. 40 2021²² to regulate teacher to be assigned as to school principals. It requires school principal

and supervisor candidates to complete PGP. In 2022, the policy was expanded with Permendikbud No.26 2022 Article 13²³, whereby PGP certification also provide opportunities to become school supervisor or other leadership positions in the education sector.

Program Design, Delivery, and Curriculum

This program aims to develop a new generation of education leaders. Ultimately, the goal is to prepare teachers to become transformational leaders in the education sector, becoming school principals, supervisors, teacher professional development instructors²⁴, thus contributing to an increase in the number of students with a Pancasila profile; or those who think critically, creatively, independently, have faith, fear God Almighty, are of noble character, work together and are globally competent.

The PGP program combines online learning using a learning management system (LMS), monthly face-to-face workshops, and individual mentoring in schools

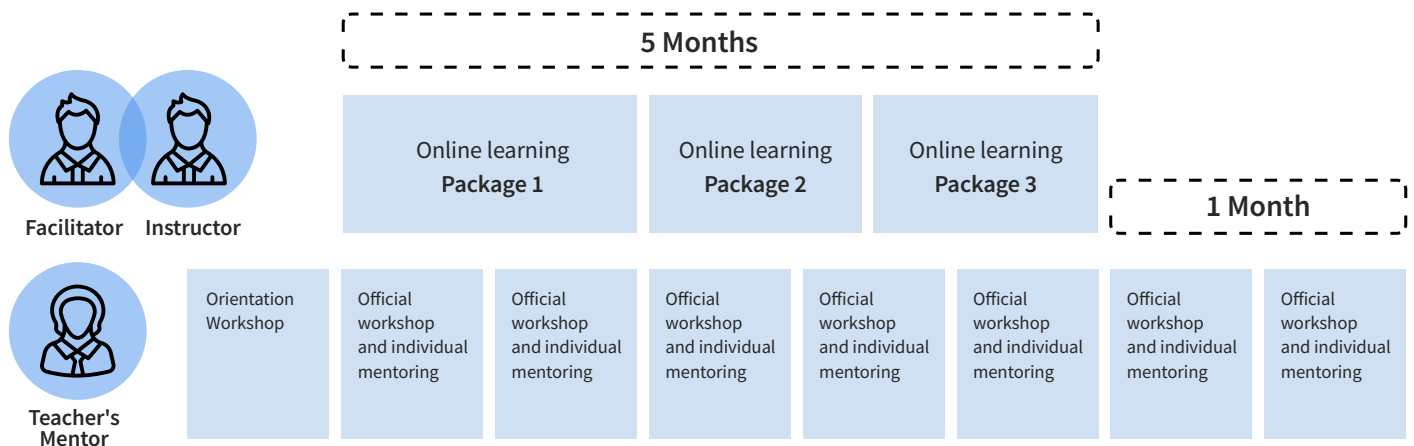
²⁰ MoECRT, 2020

²¹ MoECRT, 2023

²² Permendikbud No. 40, 2021

²³ Permendikbud No. 26, 2022

²⁴ GTK, 2023

Figure 1. *Pendidikan Guru Penggerak – Program Design*

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

over a six-month period. PGP is based on an andragogical approach throughout a six-month learning cycle, providing teachers with access to a learning community where they can collaborate and implement student centered learning projects at school during and after training. In the process, they regularly reflect on their learning, exchange feedback among participants, mentors, and facilitators, and create action plans to improve their skills. Communities of practice are developed in a natural and organic manner. The proportion of activities consists of 70 percent studying in place of work (on-the-job training), 20 percent study with peer teachers in the PGP program, and 10 percent training with facilitators, instructors and the teacher's mentor. The learning journey (see figure below) begins with an orientation workshop, where teachers are introduced to the program's policy, culture and learning targets. Subsequently, teachers explore the modules on the LMS, enabling them to acquire knowledge and skills at their own pace. In addition, teachers participate in offline workshops to deepen their understanding of the content and engage in collaborative activities with their peers (fellow Guru Penggerak participants). Throughout the program, mentors assist and provide feedback for teachers as they apply their newly acquired skills in their schools.

The online learning component is conducted through MoECRT's Learning Management System (LMS). The LMS adopts an appreciative inquiry approach, inspired by the Teacher Inquiry Program at the British Columbia Teachers Federation. Each module follows the "MERDEKA" cycle, which consists of seven steps. It begins with self-reflection on the material, followed by independent study by the participants. This is supplemented by a discussion forum and facilitated group assignments. Participants then engage in mid-learning reflection and proceed to work on individual tasks. An enrichment session with the instructor follows, where participants deepen their understanding of the topics. Subsequently, participants integrate their acquired

knowledge and develop a plan for action, known as "Aksi Nyata" for implementation in the participants' classrooms or schools as a practical application of their learning. Participants go through a total of 10 modules, organized into three categories or module packages, as follows:

- **Package 1** focuses on the vision of Guru Penggerak and the importance of creating a student-centered learning environment, highlighting the values, roles, and vision necessary to cultivate a positive learning culture.
- **Package 2** centers around student-based learning practices, encompassing inclusive and differentiated learning paradigms, mapping student learning needs, and implementing strategies for differentiation. It emphasizes socio-emotional learning and coaching as a supportive system.
- **Package 3** explores the role of instructional leaders in school management, covering decision-making, asset management within the school ecosystem, and managing programs that promote student leadership and community engagement. These modules equip teachers with the mindset and skills to prioritize student-centered learning, foster collaboration among peers, and implement asset-based school programs collectively.

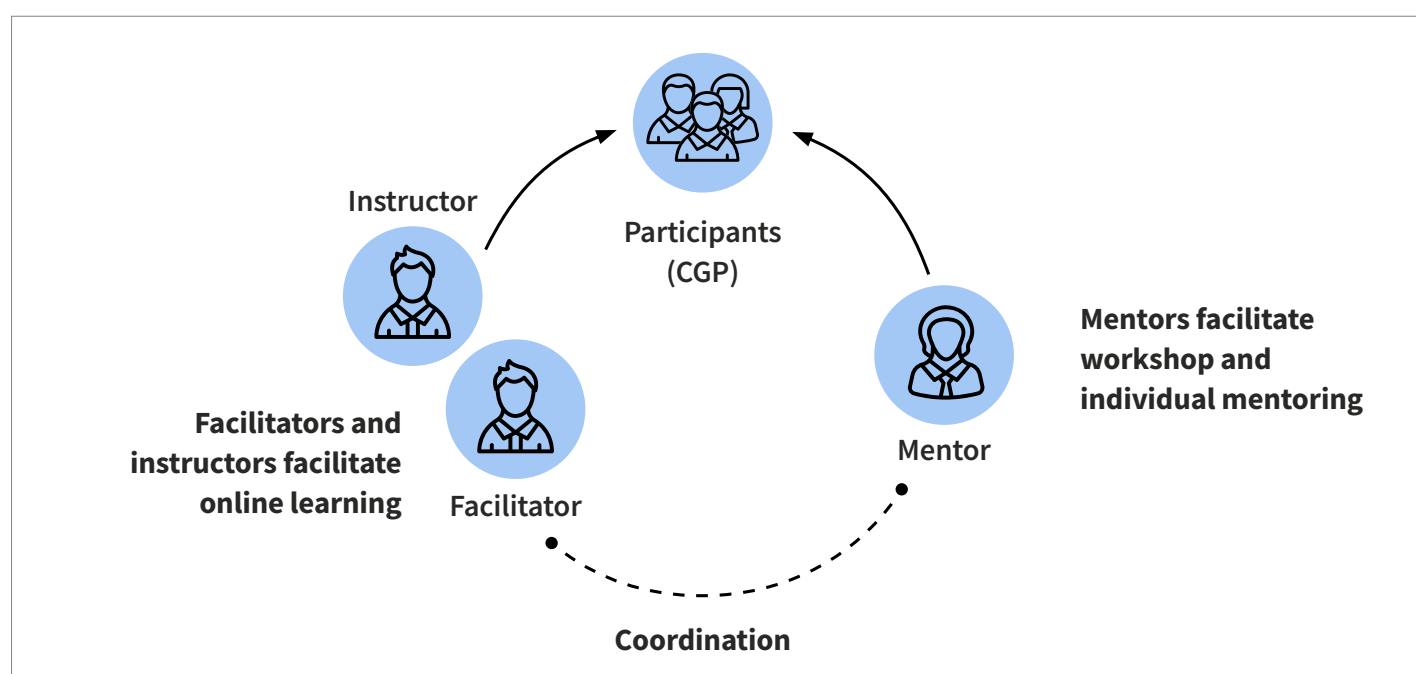
The workshops are conducted in-person at the district level, facilitated by mentors. This is an opportunity for participants to collaborate, deepening their understanding, and reflect on their teaching practices through eight six-hour workshop sessions. These include orientation followed by self-development and networking with fellow educators, the school's vision and mission based on student-centered philosophy, and learning practice in the classroom. Next, participants are expected to be able to carry out academic

supervision applying a coaching mindset. Later sessions focus on developing an asset-based and collaborative program work plan for student-centered school development and good practice sharing with school principals, offices, or learning communities.

Individual mentoring sessions take place at each participant's school. Mentors advise 3-8 PGP participants, each receiving six individual mentoring sessions with 4-training hours. Mentors observe the application of knowledge from the LMS and workshops in the classroom, provide feedback on teaching practices, support improvement plans and assess the learning process in the classroom and peer teacher coaching.

Facilitators play the role of assisting participants in the online learning process within the Learning Management System (LMS). This includes self-paced asynchronous learning and synchronous discussions. They are also responsible for assessing and providing feedback on participants' assignments. **Instructors** are tasked with enhancing the understanding of each module studied by the CGP in virtual (synchronous) environments. Lastly, **Mentors** guide face-to-face workshops and offer individual support within school settings. Each role carries specific responsibilities that are interconnected to ensure the effectiveness of the program (please see Appendix 1.D.).

Figure 2. PGP Actors roles and relationships



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

Selection process

The program is highly selective and centralized, handled by the Directorate General of Teachers and Education Personnel GTK) of MoECRT. Both civil servant (PNS) and non-civil servant (non-PNS) teachers in both public and private schools, who have at least a bachelor's degree and five years of teaching experience, are eligible to apply if they're teaching in regular schools (teachers at MoRA schools/madrasahs are not eligible). Applicants register online in the *Guru Penggerak*²⁵ website, uploading their national ID, education certificate, appointment letter, and endorsement letter from their

school principals, along with an application essay. Selection includes verification and validation of the documents and essay grading. Short-listed teachers can join the second stage of the process, where they are graded by six pedagogical competencies through micro-teaching simulations, and participate in eight leadership competency interviews, with a score ranging from one to five. The two selection criteria are:

1. The sum of their scores across the 14 different assessments should be above a certain threshold.

²⁵ sekolah.penggerak.kemdikbud.go.id

2. They should receive no score in the “less competent” category (below 3) in any of the competencies. If there are multiple candidates within a district who meet these requirements, they will then be ranked based on their overall score and selected according to the available quota. Quota for each district in each batch could be varied and will be determined by MoECRT’s discretion.

In Batch 5 in 2022, roughly 20,000 teachers were short-listed from all levels of education (kindergarten, primary, junior secondary, senior secondary, and vocational schools) in 166 districts in Indonesia. Of these teachers, roughly 8,105 were selected based on their performance on the simulation and interview.

Theory of Change

The Theory of Change for the PGP training program (see Appendix 1) aims to reach all elements of education institutions; students, schools and school personnel, creating students with a Pancasila profile. To achieve this, it establishes a student-centered system within schools, developing quality teachers, principals, and school supervisors and focuses on four final outcomes, each with a set of key indicators. These final outcomes and indicators serve as a framework for our impact evaluation study which is aimed at measuring the impact and effectiveness of the PGP program in achieving its intended goals.

Figure 3. PGP Theory of Change: Final Outcome

Final Outcome	Teachers have the moral, emotional, and spiritual maturity to behave according to the code of ethics.	Teachers plan, implement, reflect, and evaluate student-centered learning in schools.	Teachers support peers/communities in their schools to implement student-centered learning.	The teacher initiates the development of a student-centered school based on collaborative resources.
---------------	---	---	---	--

¹¹ Guru Penggerak Program aims to develop a new generation of education leaders. Ultimately, the goal is to prepare teachers to become transformational leaders in the education sector, becoming school principals, supervisors, teacher professional development instructors ¹¹



4.

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

1. 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?
2. 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

1. What are teachers' perceptions of the program?
2. 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?
3. 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.²⁶

²⁶ The counterfactual is beneficiaries' outcomes if they did not participate in the program. It is critical to have a valid measure of the counter-factual in order to estimate the causal impact.

The empirical strategy is slightly more complicated (see Appendix 4 for more details on the estimating equations). 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:

- 1. 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.
- 2. 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.
- 3. 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.

Sample Selection - Examining the Cutoff for Selection

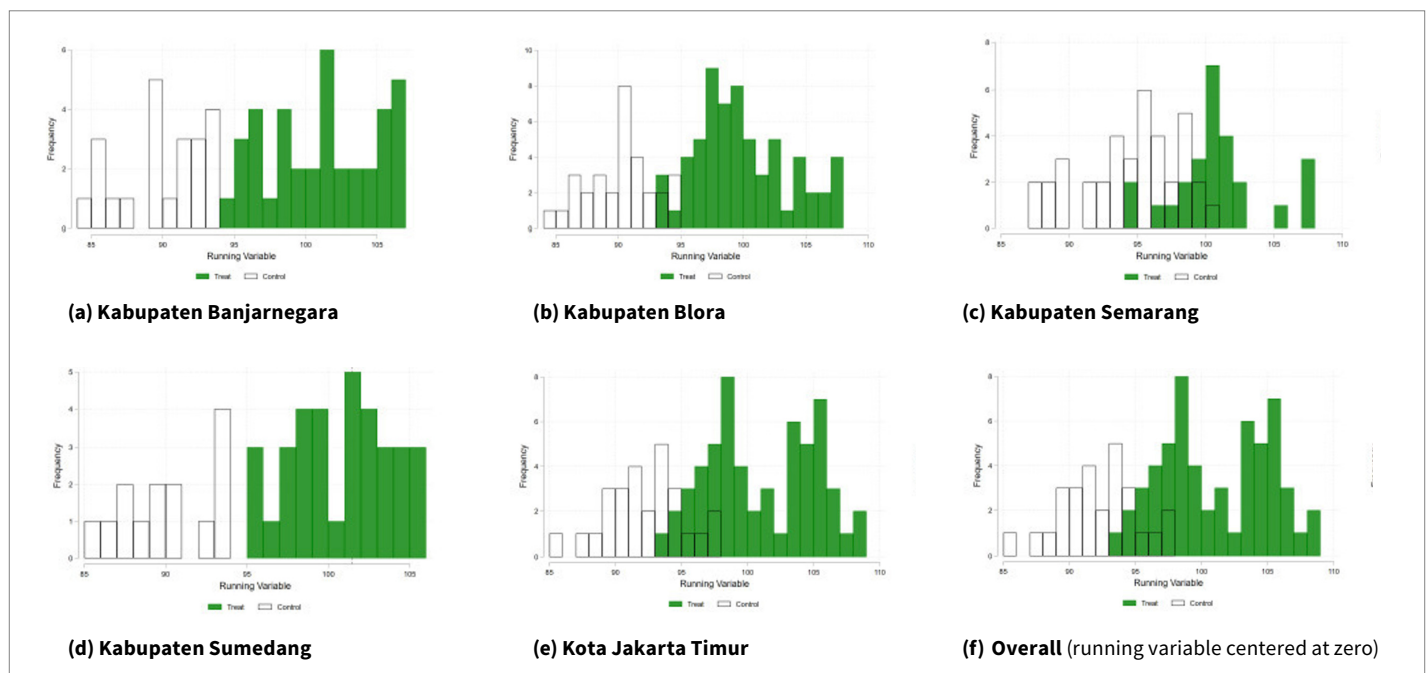
The sample is drawn from all primary-level teachers who were short-listed to participate in Batch 5. Out of these teachers, 350 teachers who were closest to the threshold for selection in their respective districts/cities, were sampled for data collection. The sampling process involved identifying five 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 as shown in Figure 4.

Figure 4. Location of Sample Districts

Figure 5 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 (see Appendix 4) 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 5(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.²⁷ 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.

Figure 5. Distribution of sampled teachers by assignment to treatment

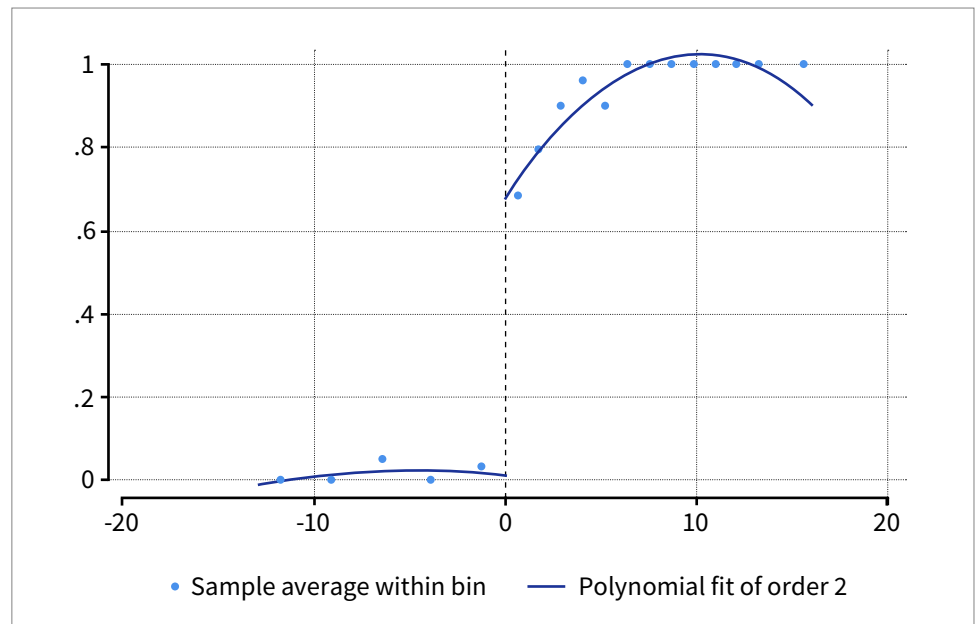
Note: The running variable is the teachers’ overall score in the second round of selection. In Figure (f) it is centered to be equal to zero at the cutoff for selection for a given teacher contract and district.

²⁷ For example, it is 94 for all teachers in Banjarnegara and 95 for all teachers in Sumedang. See the appendix on how we determine the cutoff for the districts with non-compliance. In two districts the cutoff value that maximizes the power of the first stage is slightly lower for PNS teachers.

Figure 6 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 6. 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.



Limitations

Although the internal validity of our evaluation strategy is strong, one limitation is the potential lack of generalizability of the findings to the population of PGP graduates as a whole.

First, the sample is restricted to five districts in Java. Our sampling strategy for the RDD meant that we only sampled teachers close to the cutoff, but the only districts that had a large enough sample of short-listed teachers who scored close to the cutoff were all in Java provinces. Our results cannot tell us what the impacts are in more remote and rural locations outside Java. Plausibly, the impacts would be smaller if the quality of implementation was weaker. Therefore, further research focusing in remote and rural areas is warranted to understand the impact of the program across the country.

Second, results are not generalizable to the types of teachers who scored well in the selection process, since the RDD only estimates the impact for teachers very close to the cutoff. This is a shortcoming that is inherent in any evaluation using RDD. In contrast, results of a randomized evaluation can be generalized to all treated teachers in the evaluation sample, including the best-performing teachers.

Third, the study is focused on primary school teachers, and does not include secondary school teachers. Hence, there is a possibility that the observed impact cannot be generalized to teachers of different levels.

Fourth, the study only examines the impact of teachers who participated in Batch 5. Since the program is scaling up over time—from 2800 teachers who participated in Batch 1 to 55,000 teachers who will be selected in Batch 10—the quality of implementation across regions could be more varied in later batches. The characteristics of the teachers selected in the different batches might also be different, since it was more selective at the start. Therefore, the program might have had a larger impact in the earlier batches, relative to later batches.

Another limitation is that this study has low statistical power. Regression Discontinuities Designs are more data-hungry, since analysis should be restricted to observations close to the cutoff. Moreover, the non-compliance discussed above also reduced statistical power. The implication is that we can only detect statistically significant effects that are very large.

It is also important to note that other activities related to the Merdeka programs—Sekolah Penggerak, Organisasi Penggerak—were also taking place in both the treatment and control schools. This has implications for how we interpret the results. We are measuring the treatment effect of the PGP program, relative to the status quo of other training programs that are taking place in some of the schools.

5.



DATA AND INSTRUMENTS

This study reports results based on two different data collection activities (see Figure 7 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. The data from the school visits is used to construct outcome measures to estimate the impact of the

program. We discuss each of these data collection activities in turn.

Implementation Assessment (phone survey)

The purpose of the phone survey activity was to gather teacher's and mentor's feedback on the program design and the quality of implementation. For the teacher survey, the sample was drawn from all teachers participating in PGP Batch 5 from 13 districts across Indonesia (see Table 2).²⁸

Table 2. Sample for Phone Survey and School Visit Data Collection

Phone Survey (13 districts)		School Visit (5 districts)
1. Kab. Bengkulu Utara	8. Kab. Semarang	1. Kab. Banjarnegara
2. Kab. Kudus	9. Kota Binjai	2. Kab. Blora
3. Kab. Lamongan	10. Kota Jakarta Timur	3. Kab. Semarang
4. Kab. Maros	11. Kota Salatiga	4. Kab. Sumedang
5. Kab. Pacitan	12. Kota Tanjung Pinang	5. Kota Jakarta Timur
6. Kab. Pasaman	13. Kota Yogyakarta	
7. Kab. Pekalongan		

²⁸ The sample of districts for the Implementation Assessment and Impact Evaluation is different for two reasons. First, we were able to collect data in more districts with phone surveys, whereas cost constraints meant that the in-person data collection could only take place in five districts. Second, these districts were selected for an initial design of a randomized evaluation of PGP. This design was not feasible given changes in the timing of the program. A different set of districts were more appropriate for the RDD.

Table 3. Summary of Phone Survey Data Collection

Phone Survey	Date of Survey	Sample	Measure
PGP Teacher – Round 1	June 2022	200	Quality of the LMS
PGP Teacher – Round 2	November 2022	200	Quality of workshop and individual coaching
Mentor Teacher (<i>Pengajar Praktik</i>)	December 2022	10	Mentor preparedness and implementation of workshop and coaching
PGP Teacher – Round 3	January 2023	199	Quality of overall program

Over 1,000 teachers participated in PGP Batch 5 from these 13 districts. We selected 400 teachers through a randomized sampling approach and then assigned these teachers in two cohorts of 200 teachers, conducting three rounds of data collection. The first cohort of 200 teachers was surveyed at the beginning and the end of the training. The second cohort of teachers was surveyed in the middle of the training.

For the mentor teacher survey, a total of 10 mentors were selected through random sampling. The sample was drawn from all mentor teachers in the same 13 districts of the PGP Batch 5 teachers. Since these mentors were randomly selected, we cannot link the teachers sampled for the phone survey and these sampled mentors.

The teacher phone survey was conducted in multiple rounds throughout the implementation period of PGP Batch 5. Each round aimed to collect information on different elements of the training, as summarized in the table below. The mentor survey was conducted once, during Batch 5 implementation.

The instruments used for data collection were collaboratively developed with the team from MoECRT. In each round of the survey, we collected basic demographic information such as gender, age, employment status, highest education completed, years of teaching experience, and the school level where the teachers were currently teaching.

The surveys administered to the teachers inquired about their motivation for participating in the PGP and their career

aspirations. Additionally, detailed questions were included to gauge their perception of the quality of PGP implementation. To ensure a larger sample size for data analysis, some questions from Round 1 were retained and incorporated into Round 2. Furthermore, specific questions from Round 1 were repeated in Round 3 to facilitate a comparison of the teachers' perceptions before and after participating in the program.

School Visit for Impact Evaluation

Instruments

The data collection instruments for the school visits were developed in partnership with MoECRT, with the goal of capturing most of the outcomes identified in MoECRT's theory of change (Table 4 below). Four different survey instruments were developed: a teacher survey, administered to the sampled teacher; a peer-teacher survey, administered to a teacher who teaches the same grade or an adjacent grade to the target teacher; a school principal survey; and a student survey that was administered to ten randomly selected students taught by the target teacher. We also administered two different classroom observation instruments, to measure teaching practices of the target teachers. After data collection started, an additional module for sample teachers who had become a school principal by the time of data collection.

¹¹ The data collection instruments for both phone survey and school visits were developed in partnership with MoECRT, with the goal of capturing most of the outcomes identified in MoECRT's theory of change.¹¹

Table 4. Mapping between MoECRT's groupings of final outcomes and our evaluation instruments

Major grouping	Grouping	Description	Instrument
I Teachers have the moral, emotional, and spiritual maturity to behave according to the code of ethics	I.2 Self-improvement	Teacher reflects on their own competencies and strengths and weaknesses; develops a self-development plan; and executes this plan.	Teacher survey
	I.3 Socio-emotional competencies	Teacher masters social-emotional competence in teaching practice in schools. This includes: identifying emotions, responding positively to emotions, and showing empathy to students' point of view,	Student survey
II Teachers plan, execute, reflect on, and evaluate student-centered learning in schools.	II.1 Positive discipline	Teacher (a) engages students in crafting class beliefs; (b) applies, monitors or managers positions consistently in the learning process in the classroom and school, and (c) applies the restitution process to the students when they make a mistake.	Student survey, teacher survey, Classroom observations
	II.2 Differentiated learning	Teacher frequently assesses students and adapts content and teaching practices to the needs of the child.	Student survey, teacher survey, classroom observations.
	II.3 Socio-emotional competencies	Teacher (a) structures teaching to improve students' socio-emotional skills, (b) uses various strategies to integrate social and emotional learning.	Classroom observation
III Teacher encourages peers/communities in their schools to implement student-centered learning	III.1 Create communities of learning	Teacher develops the community of learning with peer-group to enhance competency and student-centered learning	Peer survey, school principal survey
	III.2 Coaching	Teacher implements coaching to develop their peer-group to enhance student-centered learning	Peer survey, school principal survey
IV Teacher initiates student-centered school development with collaborative resource (asset) based	IV. 1 and IV.2 vision of student-centered school development	Teacher has a vision to develop student-centered learning and initiates development with relevant stakeholders	Peer survey, school principal survey

The teacher survey was administered to the target treatment and control teachers in each school. It included basic demographic characteristics, as well as questions about prior exposure to teacher training programs, and the perceived perception of the quality of the PGP, if they participated. It also asked about teachers' self-confidence as a leader, enjoyment of teaching, and whether the teacher had conducted the self-reflection activities that form part of the program: i.e., whether a teacher had done strengths mapping, had a strength mapping document, and could mention areas of improvement.

The peer teacher survey was administered to a teacher who teaches in the same grade as the target teacher, or an adjacent grade if there was no other teacher in the same grade. The main purpose was to capture a more objective measure of the target teacher's engagement with other teachers in the schools (outcome groupings III.1, III.2, and IV). It therefore included questions about whether a community of learning existed, and whether a teacher had done the following: mapped out student needs, self-reflected on their strengths and weaknesses, and discussed student-centered learning with another teacher. It also asked the teacher whether they had been observed in the classroom, and who initiated the above-mentioned activities. The enumerators then indicated whether the target teacher initiated the activity or not.

The school principal survey included the same demographic questions and questions about exposure to training as the teacher and peer teacher surveys. In addition, it included a battery of questions to measure the school principal's assessment of the quality of leadership of the sampled teacher; and questions related to outcome IV from the theory of change: whether a community of learning exists in the school, whether the school principal had been asked to reflect on the school learning process, whether someone had talked to them about the school vision. As in the peer survey, the enumerators then indicated whether the school principal mentioned the target teacher as initiating the activity or discussion, or not.

The student survey was developed to measure students' perceptions of teaching quality. It included a series of questions on a Likert scale—ranging from never (1) to always (4)—that relate to the teaching practices of the sampled teacher (for example: “the teacher asks how I am doing and how I am feeling”). It also included some questions about

their views of the teacher (e.g., “I like the way the teacher treats me when I need help), again reported on a Likert scale ranging from 1 (Strong agree) to 4 (agree), and some binary questions about specific teacher behaviors. Finally, it also asked the student to list the number of students who were afraid of the teacher or were afraid to ask the teacher questions. In order to make this salient, the student first had to list three classmates, and then report how many of these classmates were afraid of the teacher.

Two different classroom observation instruments were administered to the targeted teacher, one developed internally by MoECRT, and the World Bank Teach instrument. The benefit of using the government's own instrument is that it is more likely to capture specific teaching practices targeted by the program, since the instrument was explicitly developed with this program in mind. For example, this instrument is used as a monitoring tool by the mentors when they observe teaching, so they include indicators that are key outcomes of the program. But there is a risk that teachers are only changing behavior along the specific dimensions of this instrument. It is for this reason that we decided to also use the Teach instrument, which would provide a different and independent assessment of overall change in teaching practices. Moreover, the Teach instrument has been rigorously tested, piloted, and validated for data collection purposes in a variety of contexts, and has a detailed set of guidelines and specific training protocols for applying the instrument. We can therefore have confidence in the reliability of the measure, and data is likely to be captured with less noise.

The Teach classroom observation tool measures three broad categories of teaching practices: classroom culture, instruction, and developing socio-emotional skills. Each of these broad categories has two or four clusters of behavior (see table below), nine in total. In turn, each cluster includes a series of behaviors. Enumerators were trained to provide a score for each of these behaviors. A single measure for each cluster is calculated by taking the mean of the scores of all the behaviors related to the cluster. Then the mean of all the clusters creates a score for each category. Finally, the mean across all categories creates an overall Teach score. The overall score is standardized to have a mean of zero and standard deviation of one. The results for the overall score, as well as for each category and cluster of behaviors, are reported.

Table 5. Description of the Teach classroom observation instrument.

Category	Cluster	Behavior
Classroom culture	Supportive Learning Environment	- Teacher treats all students respectfully - Teacher uses positive language with students - Teacher responds to students' needs - Teacher does not exhibit bias and challenges stereotypes in the classroom
	Positive Behavioral Expectations	- Teacher sets clear behavioral expectations for classroom activities - Teacher acknowledges positive student behavior - Teacher redirects misbehavior and focuses on the expected behavior, rather than the undesired behavior
Instruction	Lesson Facilitation	- Teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives - Teacher explains content using multiple forms of representation - Teacher makes connections in the lesson that related to other content knowledge or students' daily lives - Teacher models by enacting or thinking aloud
	Checks for Understanding	- Teacher uses questions, prompts or other strategies to determine students' level of understanding - Teacher monitors most students during independent/group work - Teacher adjusts teaching to the level of students
	Feedback	- Teacher provides specific comments or prompts that help clarify students' misunderstandings - Teacher provides specific comments or prompts that help identify students' successes
	Critical thinking	- Teacher asks open-ended questions - Teacher provides thinking tasks - Students ask open-ended questions or perform thinking tasks
Socio-emotional skills	Autonomy	- Teacher provides students with choices - Teacher provides students with opportunities to take on roles in the classroom - Students volunteer to participate in the classroom
	Perseverance	- Teacher acknowledges students' efforts - Teacher has a positive attitude towards students' challenges - Teacher encourages goal setting
	Social and collaborative skills	- Teacher promotes students' collaboration through peer interaction - Teacher promotes students' interpersonal skills - Students collaborate with one another through peer interaction

MoECRT's classroom observation instrument has 13 different teaching behaviors. For each of these behaviors the enumerator provides a score ranging from one to four. Unlike the Teach instrument, this instrument does not have standardized guidelines or training protocols on how to score the teachers along the different dimensions. These behaviors can roughly be classified into the three elements of student-centered teaching: positive discipline, differentiated instruction, and socio-emotional skills.

We construct binary variables equal to one if a teacher received a top score for each respective teacher behavior. These binary indicators are used in the analysis rather than the Likert scales, for two reasons. First, they relate to the specific teacher competencies as expressed in theory of change. Second, it aids interpretation of results. Moreover, for most behaviors the plurality of teachers received the top score, so the meaningful variation is captured in the binary indicator. Results do not change in any meaningful way when using the Likert scales rather than binary variables in the analysis.

Table 6 below provides a summary of the MoERCT indicators, grouped by domain. The second column describes the teacher competency required to receive a top score of four. Note that these are extremely high (perhaps unrealistically high) standards for top performance.

Table 6. Summary of the MoECRT classroom observation instrument

Variable name	Teacher competency reached
Panel A. Positive discipline	
Teacher communicates reciprocally	Teacher communicates reciprocally with their students (teacher-student) and also gives students the opportunity to interact with each other (students).
Manager position	Teacher uses the principle of restitution whereby reasonable mistakes occur, and they accept the mistakes of students. The teacher asks the students why they made a mistake and helps them find a way out. This type of teacher sometimes uses the position of monitor and manager in students in certain situations. Example: "What are you going to do now? What can we do to correct this mistake?"
Motivation	Teacher fosters students' desire to learn by making during learning sessions engaging and interesting. Example: The teacher designs learning that matches the interests of the students.
Panel B. Differentiated learning	
Adjust learning environment	Teacher modifies the learning environment according to learning objectives, student characteristics, and social and emotional learning objectives.
Support and independent assignments	Teacher supports students according to their learning needs. Students can do their assignments independently as a result of the support
Modify learning strategies	Teacher smoothly modifies strategies, materials, and groupings to optimize students' opportunities to learn and meet their learning needs. Students are actively involved throughout learning and focused on work that not only develops skills and challenges them to solve problems
Concrete, clear, constructive feedback	Teacher provides constructive feedback (clear and containing development advice) not only on the student's work but also their learning process (e.g., student thinks, how the pupil solves problems, communication techniques).
Differentiated assessment	As learning progresses, an assessment process occurs that is according with the learning objectives. There is evidence of teachers' efforts to differentiate to meet students' learning needs and ensure students understand the assessment process
Panel C. Socio-emotional	
Empathetic	Teacher finds out what their student is working on and thinking. The teacher listens actively to the opinions expressed by all students
Positive emotions	Teacher displays a consistently positive expression of emotions (displays happy, enthusiastic, friendly emotions)
Socio-Emotional Content in Learning Activities	Activities that reinforce students' socio-emotional competence have been integrated into learning at: a warm opening stage, a core involving-challenging-engaging stage, and an optimistic closing stage by fusing it with academic content being studied

Construction of Indicators and Indices

We structure our results by the groupings of final outcomes, as developed by MoECRT (see Table 4 above). Within each grouping, binary indicators are constructed to indicate whether the specific derivative indicator (see Table 2) is met. We then take the arithmetic mean of all the indicators that relate to a specific grouping of outcomes (e.g., positive discipline), in order to have an index of the overall performance. The groupings related to student-centered learning have two indices, based on the student survey and classroom observations data, respectively. We deviated from MoECRT's structure, by reporting the socio-emotional competencies all in one table.

Timing of Data Collection

Data collection took place between March and April 2023 – roughly 10 months after the start of the training, and two months after the completion of the training. No primary data collection took place prior to the start of the training. During data collection the target sample of 350 teachers in five districts²⁹ were surveyed and observed in the classroom. A random sample of 10 of these teachers' students were surveyed, along with their school principals, and the peer teacher. During the implementation of the survey, 31 teachers had to be replaced because they had become school principals (the survey firm had been provided with a list of 50 potential replacement teachers). 65 percent of the school principals are female, compared to 70 percent of PGP graduates who are female. The school principals were still surveyed in order to capture basic demographic characteristics, and also their perceptions of the quality of the program and self-confidence as a leader. But, since the majority of the outcomes related to teaching practices, the additional 31 replacement teachers were also surveyed. Because there were cases where there was more than one target teacher in the same school, data were collected in a

total of 310 schools.

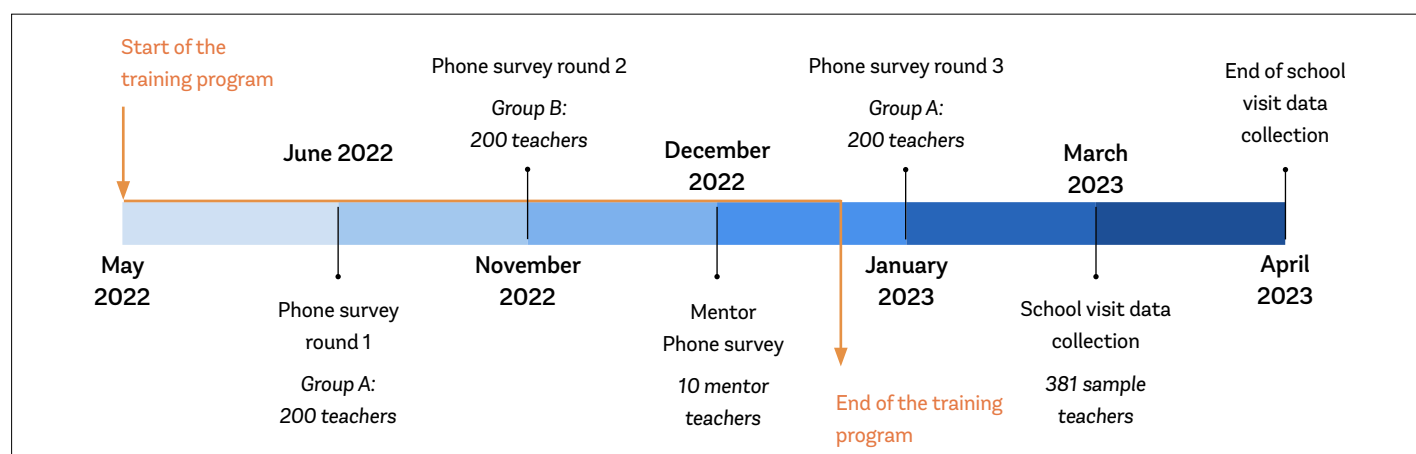
Descriptive Statistics

Table 7 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.

Figure 7. Timeline of Data Collection



²⁹ Banjarnegara, Kab. Blora, Kab. Semarang, Kab. Sumedang, dan Kota Jakarta Timur

Table 7. Teacher Characteristics - by PGP

Variable	(1) PGP Mean/SE	(2) Control Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Administrative Data				
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*
Survey data				
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.

Table A.4.1 in the Appendix shows descriptive statistics for the larger sample that includes the 31 teachers who were replaced because they became school principals. Roughly eight percent of the teachers who were originally sampled had become school principals by the time of data collection. This proportion is higher for the PGP graduates (11 percent) than the non-PGP graduates (4 percent). This is good news

from the perspective of the program, which aims to develop leadership qualities, but provides challenges from an analytical perspective, since the teaching practices of these principals cannot be observed. There is risk of selection bias if these teachers who became principals have better or worse teaching practices.

6.



RESULTS

The report first discusses the Implementation Assessment results, before moving onto the Impact Evaluation results.

Phone Survey for Implementation Assessment Results

The phone survey results are structured to answer the secondary research questions related to the quality of implementation. These include understanding teachers' perceptions of the program, evaluating program implementation quality across different districts, and assessing participants' perceptions of the policy regarding future prospects as school principals. Since it was taken three times during the implementation of the training

program, the survey provides valuable insights and regular updates for the government to improve the program.

As shown in Figure 8, teachers' reported perception of the quality of the program is extremely high. When asked to rate, between 0 and 10, how the different components of the program help them understand PGP, and how much they learnt from online Learning Management System (LMS), almost all teachers gave an overall score of 10, and the mean value ranges between 9.2 to 9.47 for these different components. There is clearly social desirability bias (or politeness) at play in these very high scores. Nonetheless, these high scores are positive indicators of teachers' perceptions of the program.

Figure 8. Quality of implementation for different components of the program

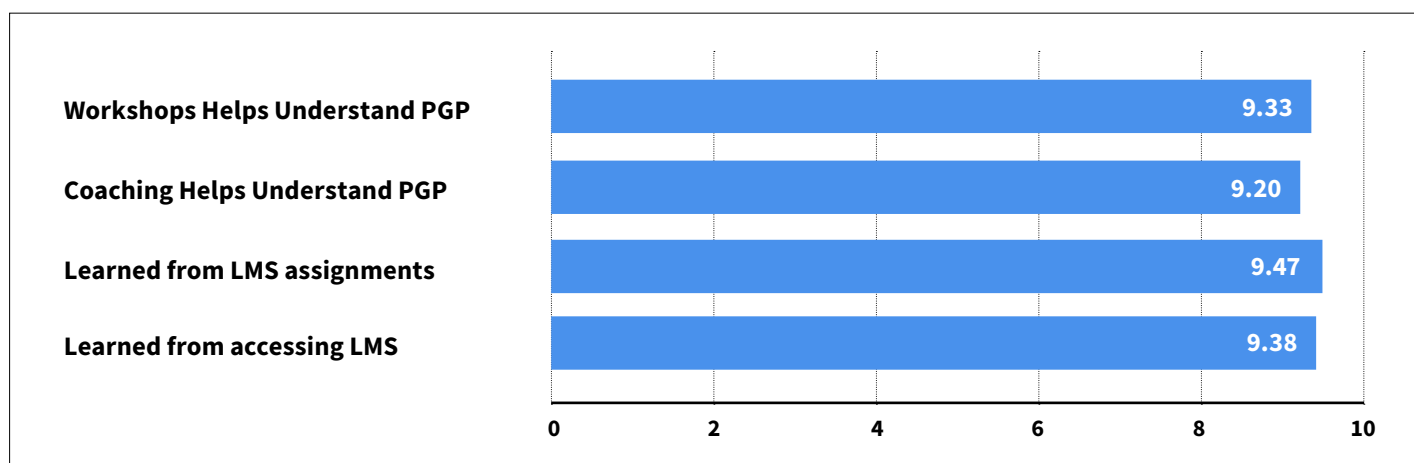
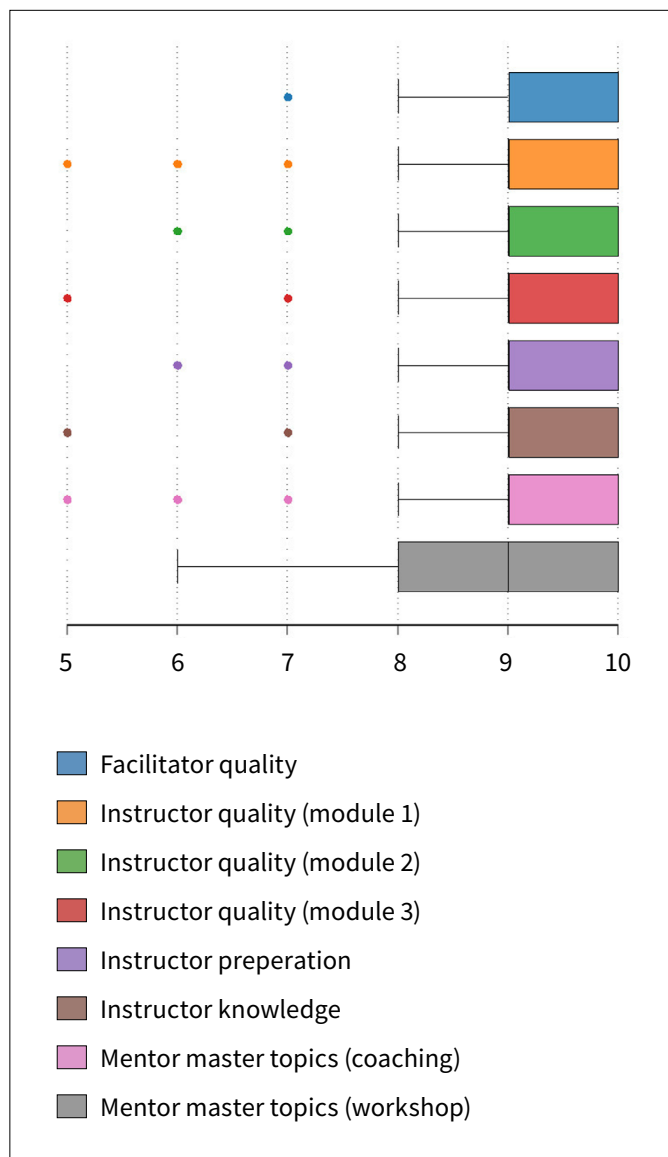


Figure 9 shows that teachers' perceptions of the quality and knowledge of the instructors and facilitators is also uniformly high. The majority of teachers provided a score of 10/10, and 75% of teachers gave a score of 9 or 10. The only exception is teachers' assessment of the mentors' mastery of topics during the workshops, where the median rating was 9/10, and a quarter of teachers gave a score below that. Although a median score of 9/10 is high in absolute terms, it is an outlier relative to all other indicators of performance.

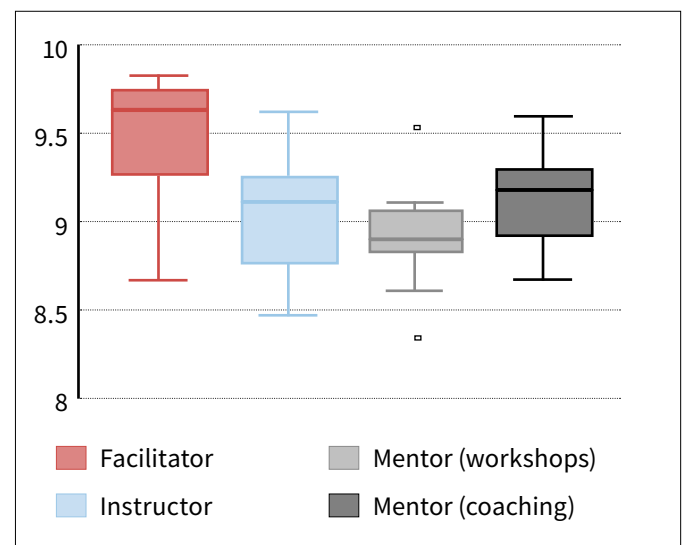
Figure 9. Distribution of perceived quality of the facilitators, instructors, and mentors



Notes. Box plots for different questions where teachers rated different domains of the quality of instructors/facilitators/mentors with a score between 0 and 10. The median score is 10 for all variables except mentor mastery of topics during the workshop, which is nine.

There is also some regional variation in the perceived quality of the different components of the programs. For example, Figure 11 shows that the average perception of the mentor's mastery of topics during the workshops ranges from 8.3 in one district and 9.6 in another district.

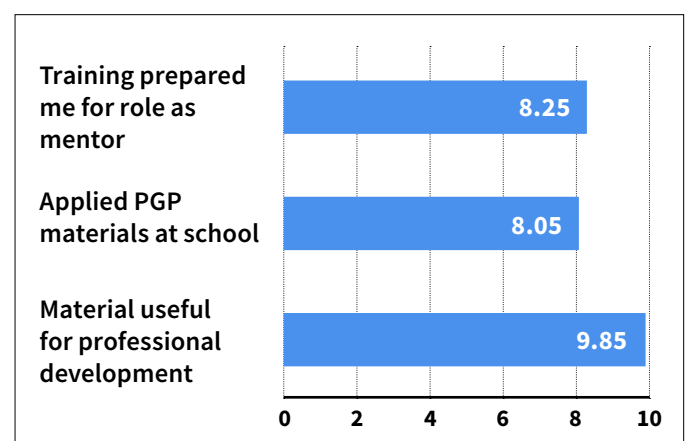
Figure 10. District-level variation in perceived quality of the facilitator, instructor and mentor



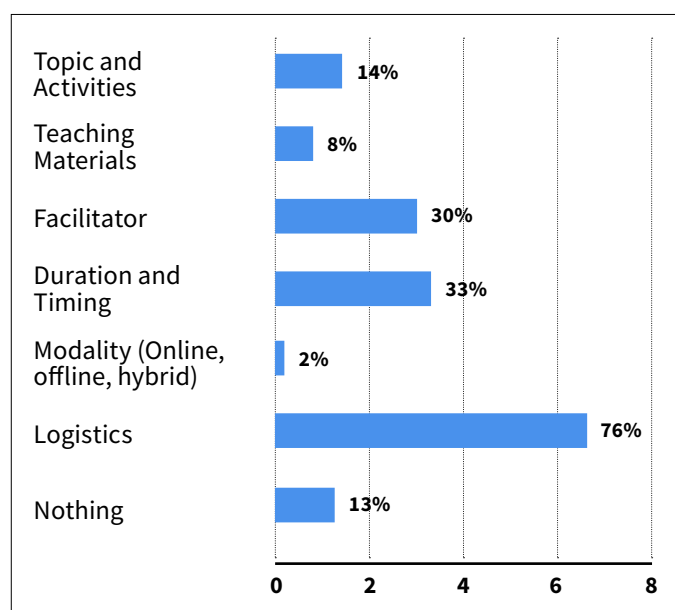
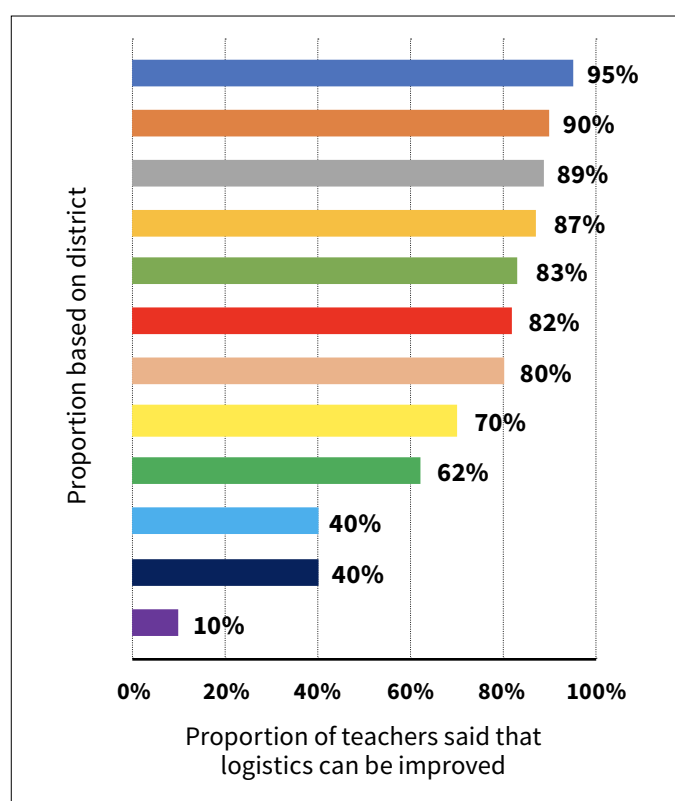
Notes. Box plots for the district average responses for different questions where teachers rated different domains of the quality of instructors/facilitators/mentors with a score between 0 and 10. Instructor is the average rating across all three modules.

We also surveyed mentors to understand mentor's training and preparation (Figure 11). Although they received handbook materials and underwent a fully online training, their rating on how much the training prepared them as a mentor is much lower than all the other indicators discussed in this section. Also, they are not uniformly applying these materials in their schools. This is concerning, since these mentor teachers should be applying these changed teaching methods, if they are expected to help other teachers do so.

Figure 11. Mentor Training Effectiveness



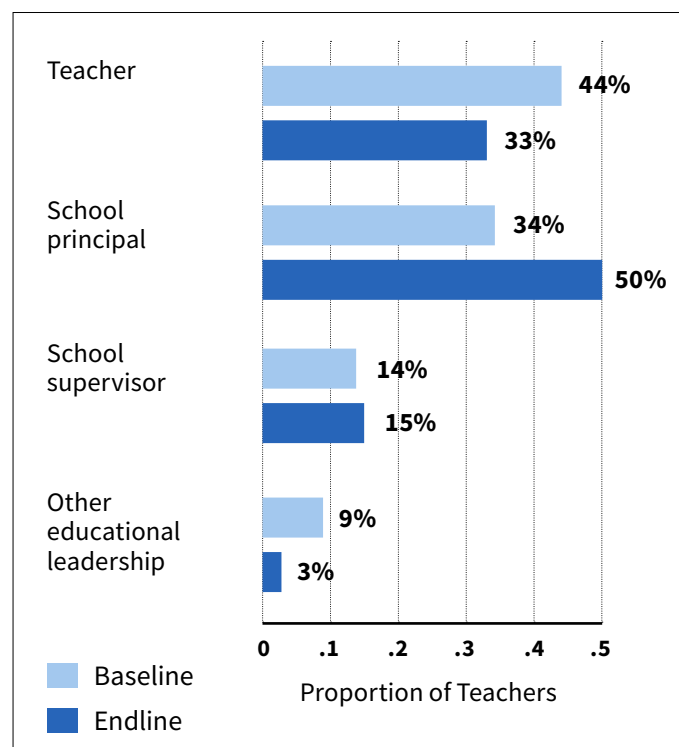
Another area for improvement is related to workshop logistics. A total of 76 percent of teachers responded that logistics needed to be improved. When asked open-ended questions about specific logistical improvements, many teachers expressed concerns about the workshop locations due to inadequate basic facilities such as functional toilets, difficulties in transportation access to reach the venues, and issues regarding meals and lack of mineral water for the entire day's sessions.

Figure 12. How workshop can be improved**Figure 13.** Proportion of teachers that said logistics can be improved

There are regional variations in the challenges faced with workshop logistics. The colors in the above figure represent different districts, indicating varying levels of agreement among teachers regarding the need for logistical improvements. For instance, in one district, 95 percent of teachers agreed that logistical improvements are necessary, while in another district, only 10 percent expressed the same sentiment. This highlights the importance of addressing the

gaps and disparities in workshop venues across different districts, as some areas have well-equipped facilities while others face challenges in this regard.

Finally, to understand teachers' perception of the new regulation³⁰ regarding their prospects of becoming school principals upon completing the PGP program, we conducted a comparison of teachers' career aspirations at the beginning (phone survey Round 1) and the end (phone survey Round 3) of the program, using the same set of 200 teachers. The findings revealed that a significant number of teachers altered their aspirations during the program. For instance, initially, 34 percent of teachers expressed a desire to become school principals, which increased to 50 percent by the end of the program. Among those teachers who did not express a desire to become school principals, the main reason cited was because they felt too young or lacking in experience. Nevertheless, this shift indicates a notable change in teachers' career goals associated with their participation in the PGP program.

Figure 14. Teachers' five-year career plans

In conclusion, the overall perception of the program among teachers is positive. The quality of implementation is generally good, although there is room for improvement such as mentor preparation and workshop logistics. Notably, a significant number of teachers have indicated a shift in their career aspirations, with more individuals expressing a willingness to take on leadership roles, such as becoming school principals and school supervisors.

³⁰ Please note that Permendikbud No. 40/2021 and Permendikbud No. 26/2022 were introduced after selection process of Batch 5. Participants of Batch 5 weren't aware of the 'benefit' of participating in the program when they applied for the program.

Impact Evaluation Results

The main results of the Impact Evaluation can be summarized as follows (see Table 8):

1

The quality of implementation was high, and teachers viewed the program highly favorably.

2

Teachers' application of positive discipline, as measured by the MoECRT classroom observation instrument, improved. However, there is no observed improvement in teachers' socio-emotional competencies, nor in the providing differentiated instruction.

3

Teaching practices, as measured using the Teach classroom observation instrument, substantially improved. The gains are especially large for the following indicators: lesson facilitation, providing a supportive learning environment, and promoting critical thinking.

4

There is evidence that PGP graduates took the initiative to support other teachers in their schools. For example, other teachers in the same school as the PGP graduates are more likely to have conducted a strength mapping exercise, been asked to self-reflect on their learning process, discussed student-centered learning, and mapped out student needs.

5

Teachers enjoy teaching more and are more self-confident in their leadership abilities. But head teachers have a lower assessment of their leadership skills, relative to the control teachers.



Table 8. Summary of main results

Indicator	Impact
ToC I. Self-development and reflection	
Self-development	Small positive
ToC II. Teaching practices	
Positive discipline	Larger positive
Differentiated learning	None
Socio-emotional competencies	None, or negative ^a
Teach instrument	Large positive
ToC III. Support other teachers^b	
Learning community	None
Self-development plan	Large positive
Coaching	Large positive
ToC IV. Initiate student-centered school development	
Vision, initiated by PGP teacher	Small positive
Other outcomes	
Enjoyment of teaching	Large positive
Self-confidence as a leader	Large positive
School principal's perception of leadership ability	Small negative

^a No impact, based on student responses, negative impact based on government classroom observations.

^b Only when the sample is restricted to schools where there are no other PGP graduates in the school

Quality of Implementation and Other Training Programs

Table 9 provides evidence of a high quality of implementation, as reported by the teachers who participated in the PGP program. Results are shown separately for each district in our sample, as well as the overall sample. The mean score for how much teachers believed they learnt from accessing the learner management system was 9.3 (out of 10). Almost all teachers indicated that they interacted with an instructor, and both the facilitator and instructor received high scores: 9.47 and 9.04, respectively. All participants reported having attended a workshop, and these were always held in person (although in some districts the workshops had a hybrid modality for a small fraction of the teachers). About 85 percent of teachers indicated that they believed that the workshops were very

relevant. Finally, all the teachers indicated that they had received mentorship in person. The average number of times that they met with the mentor was 7.6 out of 8 total number of individual mentoring sessions. The length of each session was roughly three hours. And they also rated the quality of the mentors highly, at 9.25 (out of ten). There is no strong difference across districts, with the exception of Banjarnegara, where teachers typically provided slightly lower scores in their subjective assessment of the different components of the program. We do not know whether this is because of weaker quality of implementation, or different expectations/standards from the teachers.

Overall, we can conclude that the teachers rated the quality of the program highly, and that while it is intensive (it requires a lot of time from teachers), it was well-implemented.

Table 9. Quality of Implementation by District

	Banjarnegara	Blora	Semarang	Sumedang	Jakarta Timur	Total
<i>Device used</i>						
Phone	0.95	0.95	0.88	0.95	0.96	0.95
Tablet	0.00	0.00	0.12	0.09	0.16	0.09
Laptop	0.98	0.98	1.00	0.95	1.00	0.99
Computer	0.09	0.09	0.31	0.16	0.33	0.21
<i>Device used most frequently</i>						
Phone	0.11	0.04	0.27	0.30	0.12	0.14
Tablet	0.00	0.00	0.00	0.00	0.00	0.00
Laptop	0.86	0.95	0.73	0.63	0.88	0.84
Computer	0.00	0.00	0.00	0.07	0.00	0.01
<i>Learning Management System</i>						
Learnt from LMS (0-10)	8.73	9.41	9.04	9.72	9.51	9.32
Interacted with instructor	1.00	0.96	1.00	0.95	0.98	0.98
Quality of facilitator (0-10)	9.05	9.57	9.42	9.63	9.58	9.47
Quality of instructor (0-10)	8.75	9.03	9.12	9.09	9.28	9.06
<i>Workshop</i>						
In-person	1.00	1.00	1.00	0.98	0.98	0.99
In-person and online	0.00	0.00	0.00	0.02	0.02	0.01
No. workshops attended	7.30	7.18	7.27	7.14	7.39	7.25
Very relevant	0.73	0.86	0.92	0.88	0.88	0.85
Quite relevant	0.25	0.14	0.08	0.12	0.12	0.14
<i>Mentorship</i>						
At teacher's school	1.00	1.00	1.00	1.00	1.00	1.00
At district	0.00	0.00	0.00	0.02	0.00	0.00
At another PGP school	0.00	0.00	0.00	0.00	0.02	0.00
Mentored in person	1.00	1.00	1.00	1.00	1.00	1.00
Mentored in group	0.00	0.03	0.04	0.07	0.09	0.05
Mentored online	0.00	0.19	0.00	0.02	0.11	0.09
No. times met	7.05	7.15	6.77	8.02	8.67	7.60
Length of session (in minutes)	184.09	180.27	195.77	191.06	201.05	189.39
Quality of Pengajar Praktik (0-10)	9.05	9.24	9.27	9.44	9.28	9.25
Observations	60	94	64	48	84	

Notes. The first five columns report mean values for each respective district, the final column shows mean values for the whole sample of 350 teachers. All indicators are binary variables, unless otherwise indicated. A value of 1 for a binary variable means that 100 percent of teachers said yes.

We also asked teachers in our full sample about their exposure to training in general (results in Table 10, split between the teachers who were selected to participated in Batch 5 of PGP, and those who were not). There are a few points worth noting. First, almost all teachers who were *not* selected (our control group) also participated in some kind of training over the past 12 months. So, there is a high degree of exposure to training for teachers. Second, even though the control teachers are not PGP graduates, about 40 percent of teachers in the control listed that at least one other teacher (or the school principal) in the school is a PGP graduate. *This suggests that our estimates of the school-wide impact of the program might be substantially under-stated, because so many control teachers have already been influenced by peer teachers who graduated from PGP in previous batches.* Third, a higher proportion of the control teachers indicated that their school participated in pilot *Kurikulum Merdeka* (KM): 43 percent, relative to 33 percent in the treatment group. If teachers improve their teaching practices and leadership skills because of the other programs operating in schools—*Organisasi Penggerak*, *Sekolah Penggerak*, and KM—then

our estimates might underestimate the true impact of the PGP program.

As for the characteristics of the training that they attended (Panel B), only about 70 percent of Batch 5 teachers indicated that they received training in PGP. This might be because the survey allowed them to only select one training, and they had participated in another training more recently. This again highlights the high degree of teachers' exposure to training. Moreover, teachers in the treatment group are generally more satisfied with the PGP training that they participated in, relative to the control, and they are substantially more likely to believe that the training will greatly improve their teaching. This is encouraging for the program. Finally, a high proportion of teachers who did not participate in the program have learnt about the different components covered in PGP, such as student-centered learning (97 percent), differentiation (89 percent), and the main assets/capital (73 percent). Nevertheless, these proportions are larger, and close to 100 percent, for Batch 5 teachers.

Table 10. Exposure to training (by selection into Batch 5)

Variable	(1) Selected batch (5) Mean/SE	(2) Control Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Panel A. All teachers				
Participated in training	100.243 (0.258)	0.949 (0.019)	0.979 (0.007)	0.047***
Teacher or school principal in PGP	0.693 (0.031)	0.394 (0.042)	0.743 (0.022)	0.544***
School participated in Organisasi Penggerak	0.683 (0.032)	0.226 (0.036)	0.249 (0.022)	0.036
School participated in Sekolah Penggerak	0.243 (0.029)	0.073 (0.022)	0.081 (0.014)	0.013
School participated in pilot Kurikulum Merdeka	0.050 (0.015)	0.431 (0.042)	0.365 (0.025)	-0.103**
		137	381	
Panel B. Teachers who participated in training				
Training in PGP	0.704 (0.029)	0.000 (0.000)	0.458 (0.026)	0.704
Utilization of Merdeka Mengajar Platform	0.095 (0.019)	0.369 (0.042)	0.190 (0.020)	-0.275***
Satisfied with training (0-10)	9.070 (0.063)	8.377 (0.106)	8.828 (0.058)	0.693***
Greatly improve teaching	0.856 (0.023)	0.531 (0.044)	0.743 (0.023)	0.325***
Learn about student-centered learning	1.000 (0.000)	0.969 (0.015)	0.989 (0.005)	0.031***
Learn about restitution process	0.996 (0.004)	0.792 (0.036)	0.925 (0.014)	0.204***
Learn about differentiation	0.996 (0.004)	0.885 (0.028)	0.957 (0.011)	0.111***
Learn about main assets/capital	1000 (0.000)	0.731 (0.039)	0.906 (0.015)	0.269***
Know the main assets/capital	1000 (0.000)	0.800 (0.035)	0.930 (0.013)	0.200***
N	243	130	373	

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.

Self-reflecting and Self-improving Teachers

The first set of outcomes relates to teachers' ability to self-reflect and improve based on their own perceived strengths and weaknesses, as reported in Table 11 and Figure 15. The first column in Table 11 shows the mean value of the outcome for teachers who were not selected but scored one or two points below the cutoff value for selection in their district. We can interpret this as the counterfactual outcome: the mean in the control. Column (2) shows the impact of the program when the sample is restricted to teachers who only scored two points above or two points below the cutoff for selection in their districts. This is our preferred estimate of the treatment effect. For robustness, column (3) shows the impact when the sample is restricted to teachers who scored three points

below or three points above the cutoff; and column (4) shows results when using the full sample of teachers but allowing for a quadratic relationship between the running variable and the dependent variable.

We only confidently conclude that there is a positive impact if the results are consistently large in magnitude across three different estimation strategies in columns (2) to (4)

For exposition purposes, Figure 15 shows the results of Table 11 graphically, when the sample is restricted to those who scored two below or two above the cutoff. The blue bar indicates the mean in the control group (from column (2)), and the orange bar shows the estimated treatment effect (column (3)).

Table 11. TOC I. Self-reflection and self-development

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.70 (0.03)	0.14 (0.13)	0.03 (0.12)	-0.04 (0.09)
Maps strength	0.90 (0.06)	-0.12 (0.22)	-0.16 (0.21)	0.01 (0.16)
Shows strength mapping document	0.20 (0.07)	0.56** (0.28)	0.29 (0.24)	-0.06 (0.18)
Can mention area for improvement	1.00 (.)	-0.03 (0.05)	-0.05 (0.05)	-0.07 (0.05)
Observations	31	89	127	350

Notes. Row headings denote the dependent variable, x refers to running variable, centered at zero. The first column is the mean for non-PGP teachers who scored one or two points below the cutoff for selection their district. Columns (2) to (4) show the local average treatment effects, estimated using equations (1) and (2). In column (2) the sample is restricted to teachers who scored between two points below and two points above the cutoff. In column (3) the sample is expanded to teachers who scored between three below and three above the cutoff. In column (4) the same is expanded to all surveyed teachers. All estimations control for the teacher's contract type and x , interacted with treatment. Column (4) also controls for 2, interacted with treatment. Each regression includes additional control variables that are selected using the adaptive Lasso procedure. Baseline variables that are either predictive of treatment or predictive of the dependent variable are included. Data from the teacher survey. All dependent variables, except for the indices, are binary. "Index" is the arithmetic mean of the all the other variables in the panel. "Map strength"=1 if a teacher mentioned that they have mapped their strengths and weaknesses. "Shows strength mapping document"=1 if a teacher can show the document. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 15. Reflection and self-development

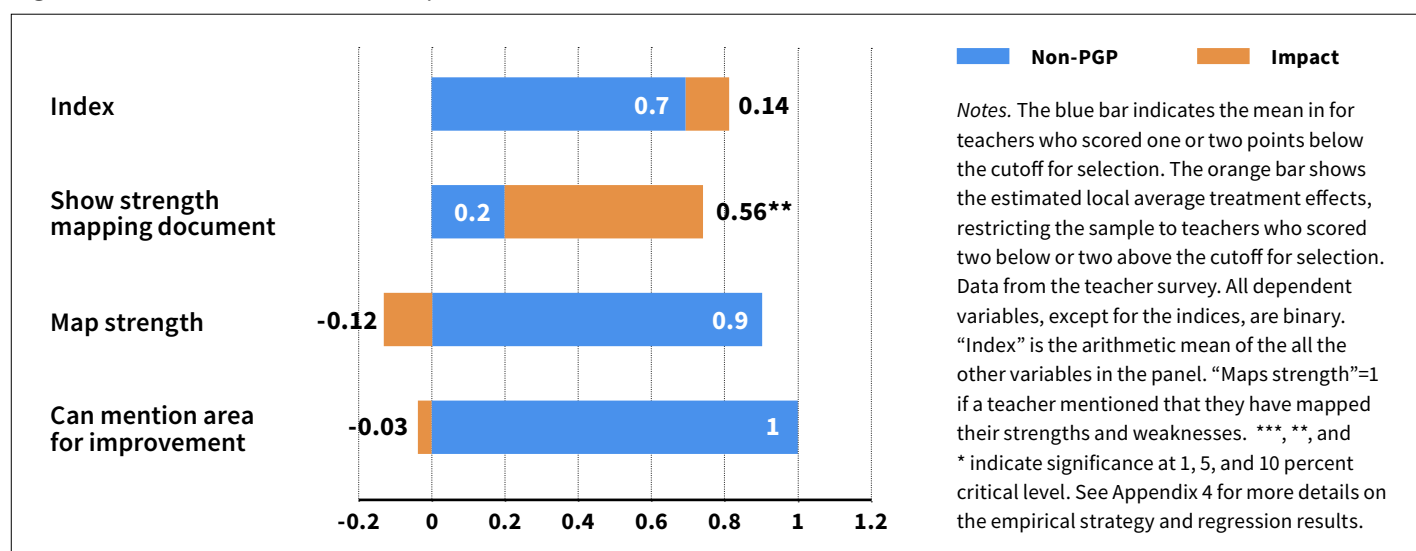


Figure 15 shows that there was no impact on the proportion of teachers who indicated that they mapped their strengths or weaknesses, mostly because all teachers, including those in the control, stated that they did so. However, there is 56 percentage point increase in the proportion of teachers who can show evidence of a mapping document. This represents an almost fourfold increase, relative to only 20 percent of teachers in the control who can show such a document. There is also no change the proportion of teachers who were able to mention areas that they want to improve on. But, again, all teachers in the control were able to mention areas of improvement, so there are substantial ceiling effects (i.e., not much “room to grow” even in the control group).

For simplicity all future analysis will only show results visually but will discuss whether the results are robust to different bandwidths and specifications.

Student-centered Teaching

The second family of indicators, according to MoECRT’s theory of change, is student-centered teaching. Student-centered teaching has three sub-components: positive discipline, differentiated instruction, and socio-emotional skills. They are defined as follows:

- 1) Positive discipline.** Teachers: (a) have clear rules for appropriate behavior and engage students in crafting these rules; (b) apply the manager position consistently in school³¹; and (c) apply the restitution process when students misbehave. Restitution means that teachers do not punish students, but rather help students understand what they did wrong.³²
- 2) Differentiated instruction.** Teachers frequently assess students and adapt the content and teaching practices to the needs of the child.
- 3) Socio-emotional skills:** Teachers structure teaching to improve students’ socio-emotional skills and use various strategies to integrate social and emotional learning.

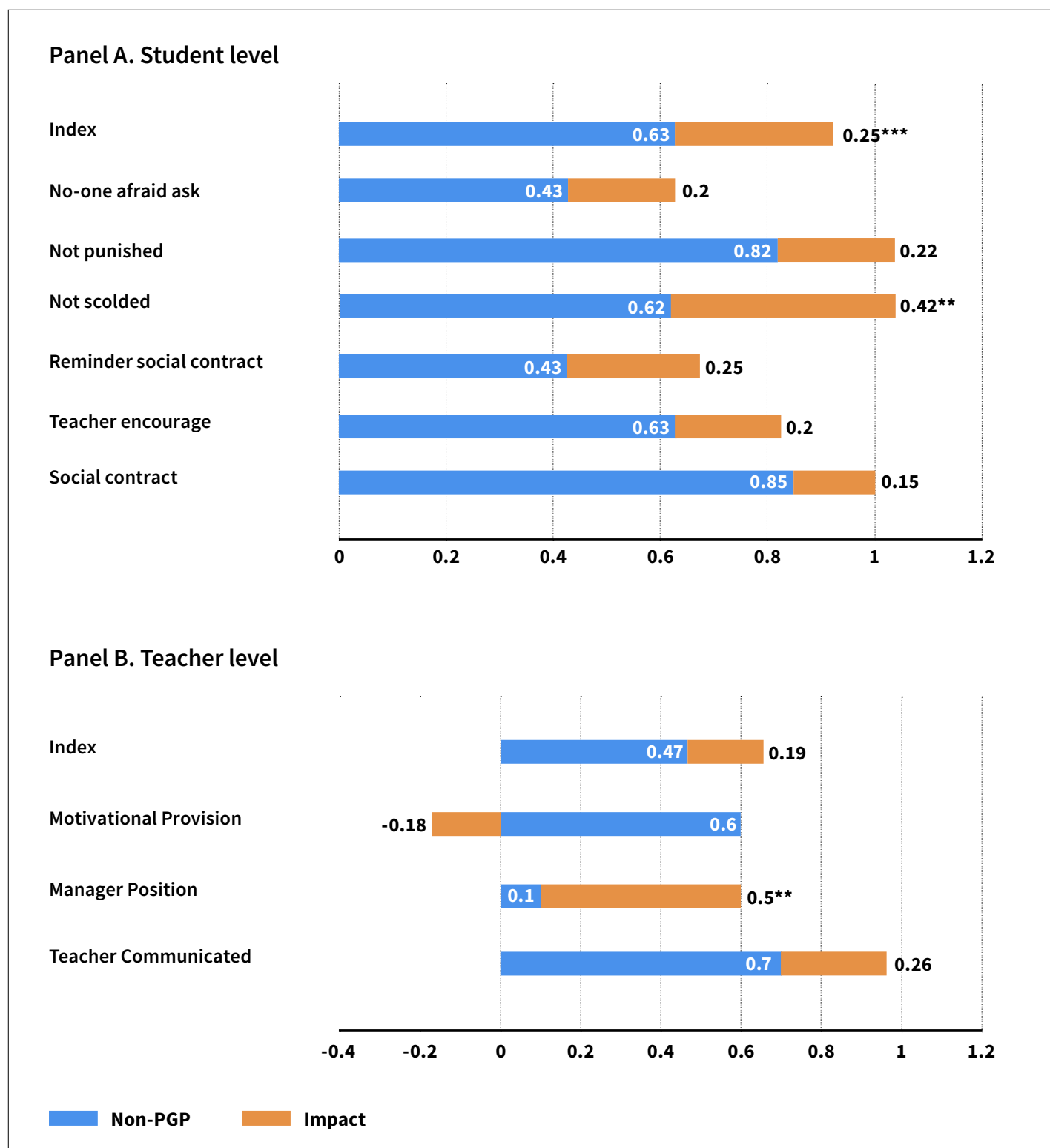
As shown in Figure 16, there is consistent evidence of improvements in application of positive discipline, based on both the student-level (Panel A) and teacher-level (Panel B) data. Students are more likely to state that their class has a social contract (or class agreement), and that the teacher reminds them of the class agreement. This is a key component of positive discipline that was emphasized in the PGP program. Students are also substantially less likely to state they have been scolded or punished by the teacher in the past, and less likely to mention someone who is afraid to ask the teacher questions. Although only one outcome is statistically significant, the impacts are large in magnitude, and consistently large across the different estimation approaches. Moreover, the impact on the overall index is large—suggesting a 40 percent improvement in the proportion of students who indicated positively to the different questions— and statistically significant at the one percent level.

These results are mirrored in the observed teaching practices, using the MoECRT instrument. PGP teachers were 50 percentage points more likely to demonstrate the “manager” position during the classroom observations, compared to only 10 percent for the non-PGP teachers. This is a sixfold increase. The manager position is also a key component of positive discipline and relates to the social contract: the teacher is required to apply the restitution process when children misbehave by reminding children of the social contract and asking them to reflect on why they made a mistake. PGP teachers are also more likely to communicate reciprocally with their students and provide opportunities for them to learn and interact from each other, although this impact is smaller in magnitude and not statistically significant. There is no evidence of improvements in “motivation provision”, which is defined as providing engaging lessons matched to the need of the child.³³ The impact on the overall index is not statistically significant, but is large in magnitude: PGP teachers are 40 percent more likely to master the different competencies related to positive discipline.

³¹ ‘The manager position was the term used by MoECRT during PGP training to indicate how teacher positioned themselves in implementing positive discipline. This position expects teachers to apply restitution process such as asking students to be responsible for their behavior, supporting students to find solutions to their problems, and analyzing their needs.

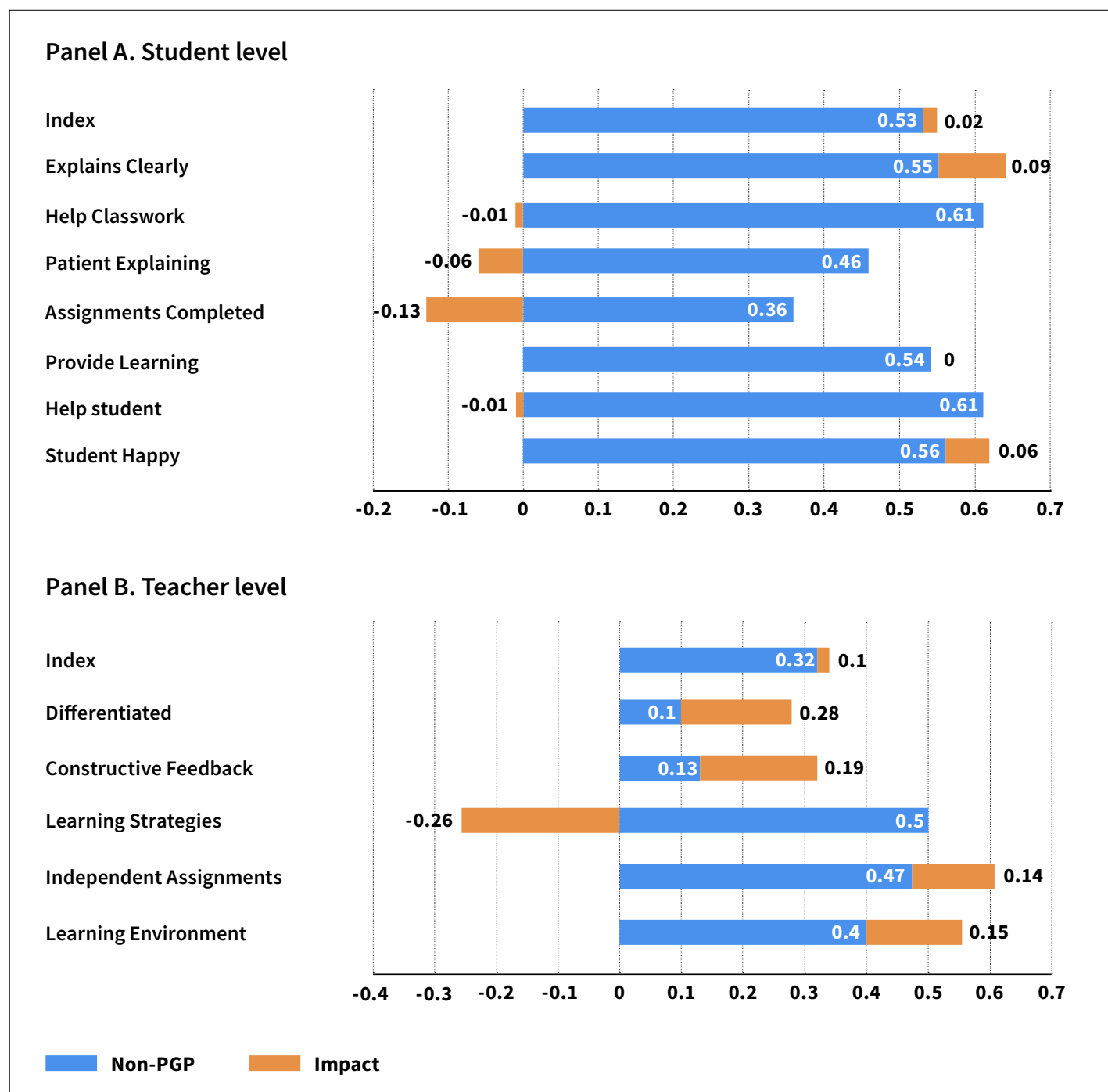
³² The classroom observation instrument defines the “manager position” as a teacher who applies restitution, so we do not draw a distinction between (b) and (c) in our analysis.

³³ One could argue that this relates more to differentiated instruction than positive discipline.

Figure 16. TOC II.1 Student-centered teaching – positive discipline

Notes. See Figure 15 notes for description of how data is constructed. Data in Panel A is at a student level; Panel B is at a teacher level. In each case the variable “Index” is the arithmetic mean of the constituent binary indicators. See Table 6 for construction of the dependent variables in Panel B.

In contrast, there is no consistent evidence for a positive impact on differentiated instruction (Figure 17). The impact on the student-level index is very small and not statistically discernable from zero. There is a larger impact on the teacher-level outcomes, but the impact on the overall index is not statistically significant. There are large increases in the proportion of teachers who provide good feedback to students and provide differentiated assessment. But a negative (albeit statistically insignificant) impact on teachers’ modification of learning strategies. Overall, we cannot conclude with a high degree of confidence that PGP teachers are more likely to enact differentiated instruction.

Figure 17. TOC II.2 Student-centered teaching – differentiated learning

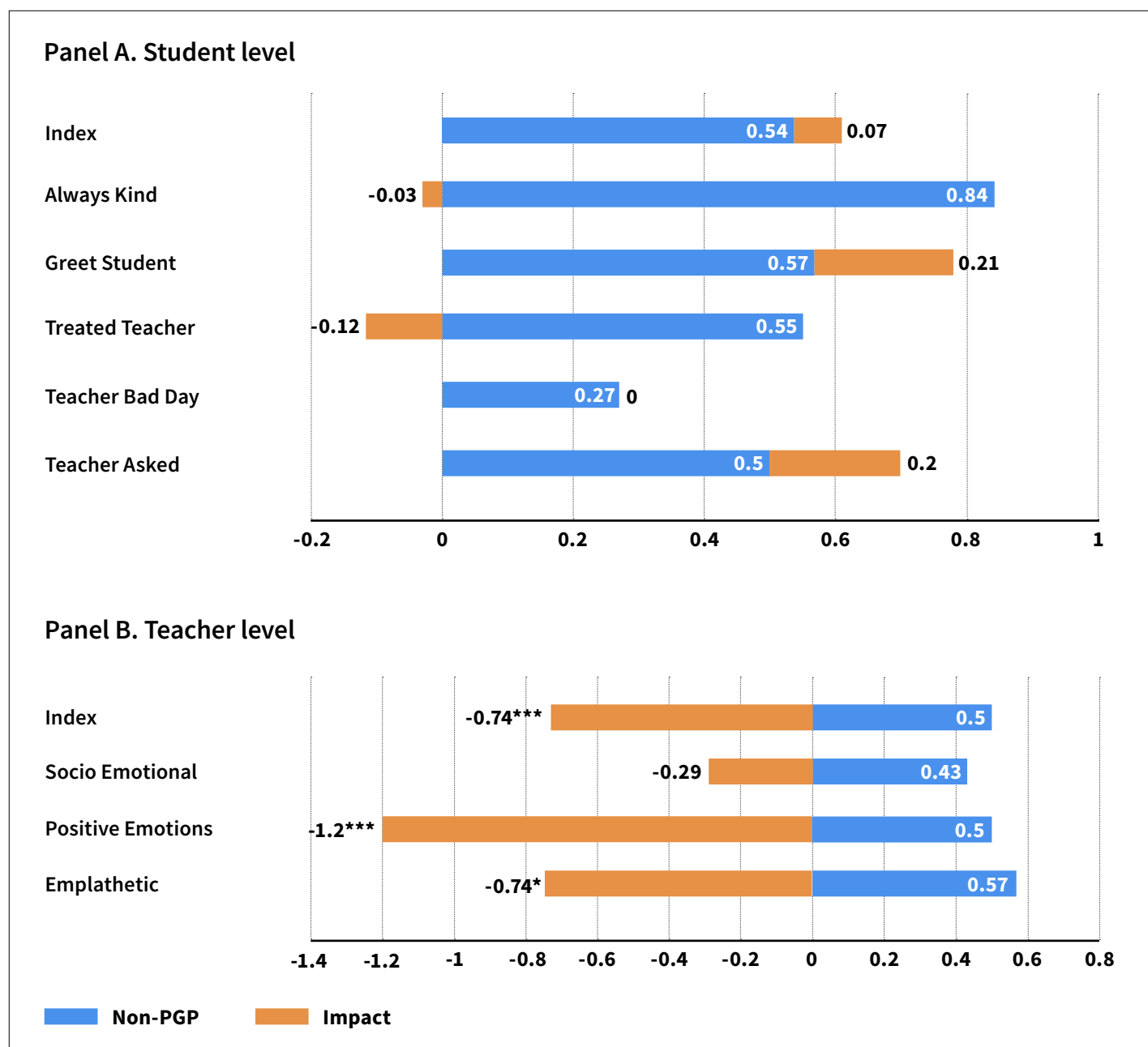
Notes. See notes in Figure 15. Data in Panel A is at a student level; Panel B is at a teacher level. In each case the variable “Index” is the arithmetic mean of the constituent binary indicators. See Table 6 for construction of the dependent variables in Panel B.

Next, we examine teachers’ socio-emotional skills in the classroom, the third dimension of student-centered teaching. Panel A in Figure 18 shows that students are no more likely to provide a positive assessment of how they are treated by the teacher. In fact, there is a reduction in the proportion of students who like the way they are treated by the teacher. Similarly, the classroom observations suggest that PGP teachers might show weaker socio-emotional skills. But these negative estimates are not consistently large in

magnitude across different specifications. Overall, we cannot conclude with any confidence that the program improved teachers’ socio-emotional skills.

This null result is consistent with the international evidence that it is difficult to improve adults’ socio-emotional skills. A recent review of evaluations of programs that aim to improve socio-emotional skills concluded that programs targeting adults had small impacts (if at all), and that programs work better for younger participants.³⁴

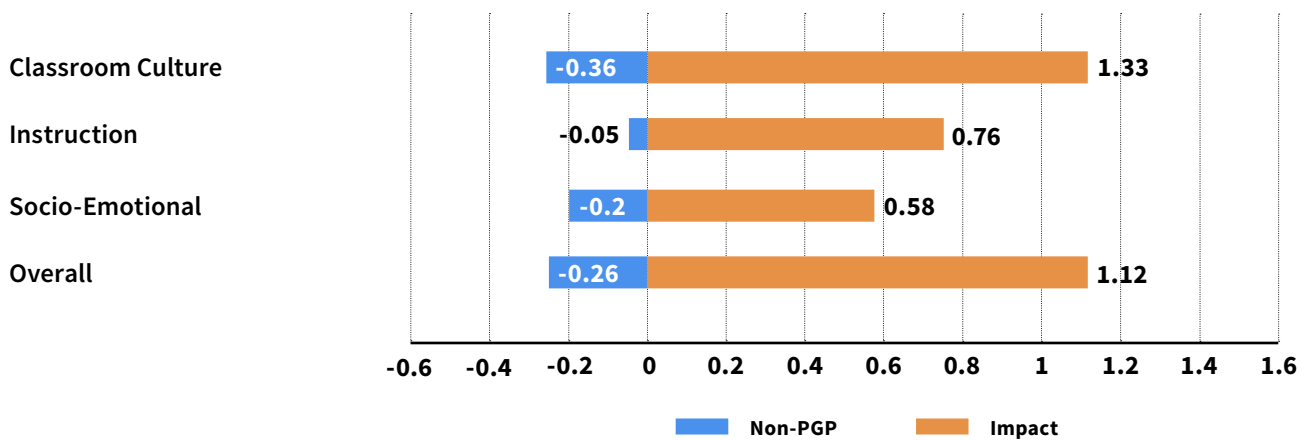
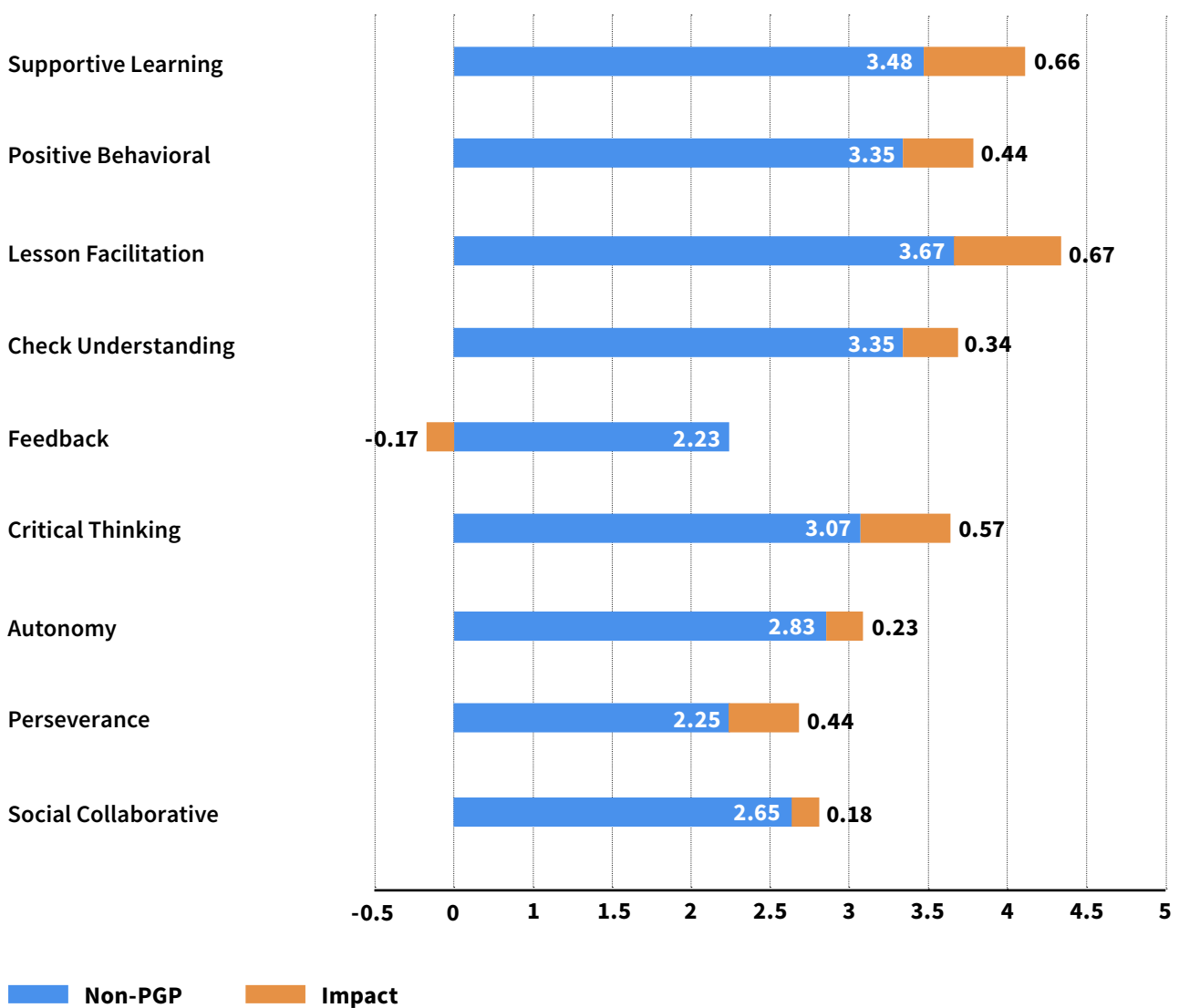
³⁴ Puerta, et al., 2016

Figure 18. TOC II.3 Student-centered teaching – socioemotional competencies

Notes. See notes in Figure 15. Data in Panel A is at a student level; Panel B is at a teacher level. In each case the variable “Index” is the arithmetic mean of the constituent binary indicators. See Table 6 for construction of the dependent variables in Panel B.

Finally, Figure 19 shows results from the classroom observations conducted using the World Bank *Teach* instrument. Teachers who participated in the PGP program improved their teaching practices (as measured by *Teach*) substantially, by over one standard deviation. As a general rule, an impact of 0.6 or larger is considered large, and an impact of 0.2 or smaller is considered small. The estimated impact on the overall score for the reduced sample—i.e., teachers who score just two points below or two points above the cutoff—is not statistically significant, but it is very large, and the estimates are statistically significant for the expanded samples, even though these effect sizes are smaller (see Appendix 4). The largest and most robust improvements are lesson facilitation, a supportive learning environment, and encouraging critical thinking.

Although there is not a perfect mapping between the two instruments, it is likely that classroom culture relates closely to positive discipline. Classroom culture includes indicators such as treating students with respect, using positive language, setting clear behavioral expectations, acknowledging positive student behavior, and redirecting misbehavior. It is important to note that “socio-emotional” is a very different meaning to what is in MoECRT’s theory of change. The *Teach* classroom observation instrument measures whether the teacher is using teaching practices that will promote a students’ non-cognitive skills; it does not measure whether teachers themselves exhibit socio-emotional competencies, which is arguably much harder to change.

Figure 19. TEACH Scores**Panel A. Indices****Panel B. Separate indicators**

Notes. See notes in Figure 15. Data from Teach classroom observations instrument. The dependent variables The indices in Panel A are arithmetic means of their constituent indicators standardized to have a mean of zero and standard deviation of one. The dependent variable Overall is the arithmetic mean of these three indices, also standardized to have a mean of zero and standard deviation of one. Effect sizes in Panel A are reported in standard deviations. The dependent variables in Panel B range between 1 and 4.

Leadership: Communities of Learning, Coaching, and School vision

For the final set of outcomes relate to the impact that PGP teachers have on other teachers in the school, and school-wide vision and leadership.³⁵ Unless otherwise stated, the results come from data collected in the peer teachers. Overall, we observe no evidence of positive effects on these school-wide outcomes. But there is a positive impact when we restrict the sample to schools where there are no other PGP teachers in the school. This suggests that the positive spillovers onto the rest of the school may have already taken place, due to previous PGP graduates in our sample of schools.

Table A.4.7 in the appendix shows that there is no evidence that the peer teachers are more likely to participate in learning communities or engage in self-development as a result of the target teachers' participation in the program. Similarly, there is no evidence on improvement in instructional leadership (Table A.4.9), or the target teacher's engagement with the school community about a school vision for student-centered teaching (Table A.4.11).

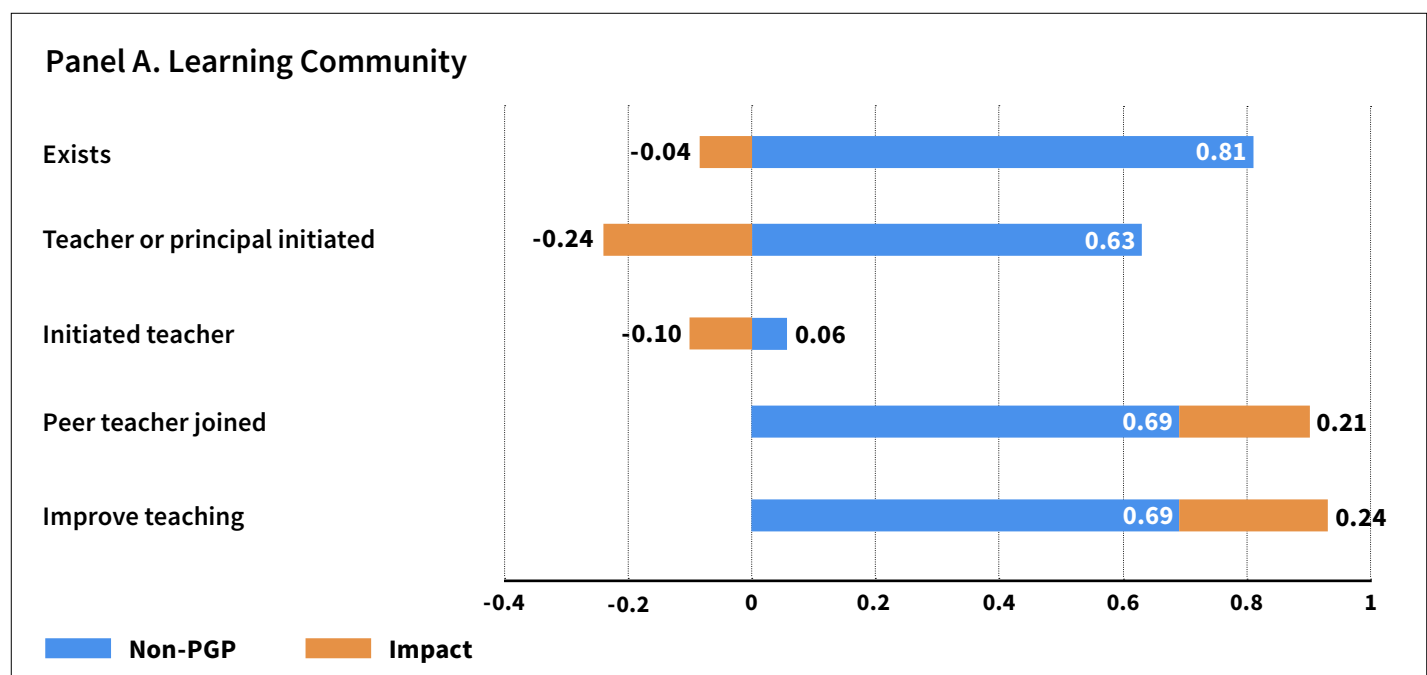
Part of the reason for the muted impact on these outcomes could be that about 40 percent of teachers in the control group mentioned that there is at least one teacher in their school who participated in PGP training. So, the peer teachers in the

control group might already have benefited from exposure to other PGP teachers. Similarly, 34 of the target teachers in the treatment group indicated that other teachers in the school or the school principal are PGP graduates. In these cases, our target teacher might not have initiated the conversations with the peer teacher about student-centered learning, because someone else had already done that.

We test for this by further restricting the sample to schools where no teacher has participated in PGP (control), or only one teacher has participated (treatment group). With this restricted sample there is much stronger evidence of instructional leadership initiated by the treated teachers.

Figure 20 shows that the peer teachers are substantially more likely to indicate that they mapped their strengths and weaknesses. They are also more likely to indicate that they were asked to self-reflect on their learning process, and more likely to mention the target teacher as the person who initiated this process.³⁶ Similarly, Figure 21 shows large improvements in the proportion of peer teachers who indicated that they were asked by a fellow teacher to discuss student-centered teaching, they were asked to map out student needs, and that this was initiated by the target teacher. There is also some evidence (Table 14) that the PGP teachers are more likely to initiate a conversation with the school principal about the school's vision for student-centered teaching.

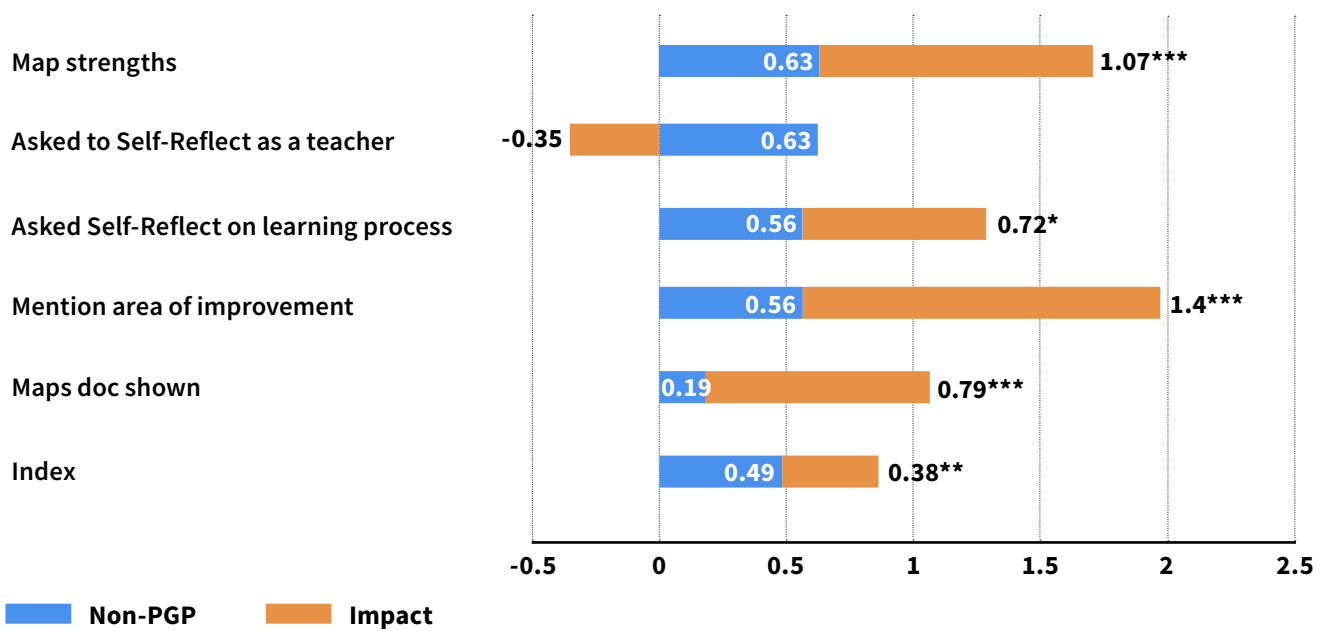
Figure 20. TOC III.1 Create communities of learning (restricted sample)



³⁵ Category III of the theory of change is: "teachers encourage peers/communities in their schools to implement student-centered learning". This is in turn divided into two sub-components: (i) establishment of communities of learning; and (ii) applying coaching principles and engaging with their peer teachers to create student-centered learning. Category IV is: teachers have a vision for student-centered school development.

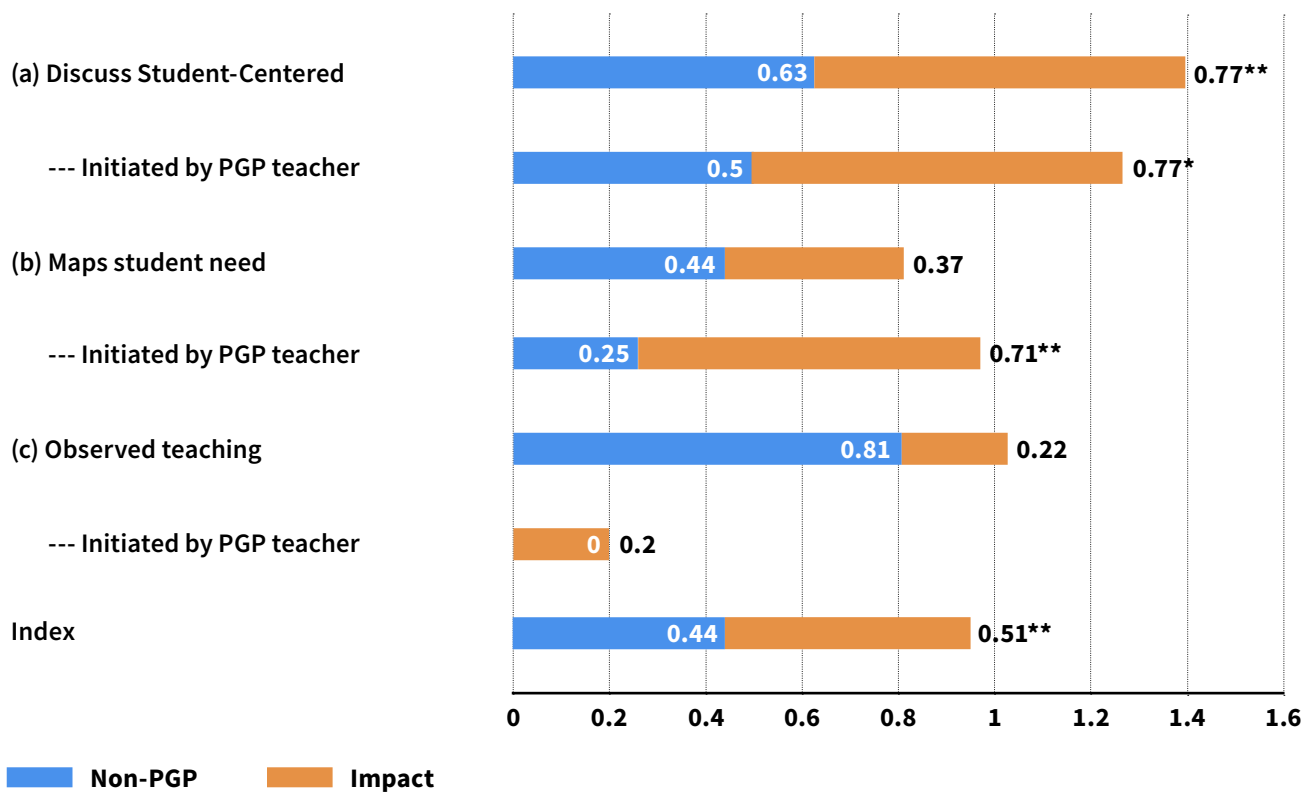
³⁶ For some specifications and samples these effect sizes are too large, since it implies that more than 100 percent of the peer teachers performed the activity in the treatment group and deserve further inspection (as done in the appendix), but what is important is that they are robust to different specifications and samples.

Panel B. Self-development plan (peer survey)

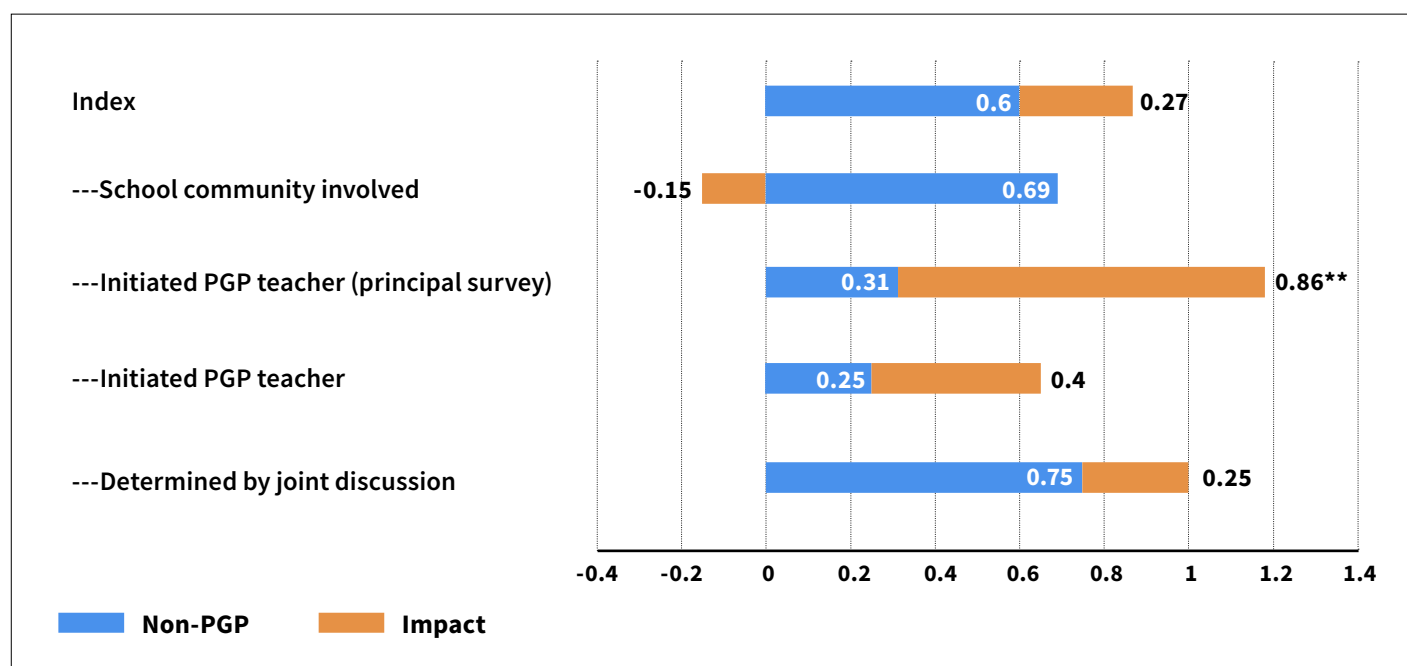


Notes. See notes in Figure 15. All dependent variables are constructed from the peer teacher survey. Index is the arithmetic mean of all the other dependent variables shown in both the panes. All variables are binary. The sample is restricted to schools where there is no other GPG graduate teaching at the school.

Figure 21. TOC III.2 Coaching (Restricted sample)



Notes. See notes in Figure 15. All dependent variables are constructed from the peer teacher survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary. The sample is restricted to schools where there is no other GPG graduate teaching at the school.

Figure 22. TOC IV.2 Vision of student-centered school development (restricted sample)

Notes. See notes in Figure 15. All dependent variables are constructed from the peer teacher survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary. The sample is restricted to schools where there is no other PGP graduate teaching at the school.

Note that the program requires participating teachers to perform some of these activities with other teachers in the school. So, it is very encouraging from a program implementation perspective that the teachers are initiating these conversations about student-centered teaching with their fellow teachers, but this does not constitute evidence of improved leadership skills overall. We also do not see any impact on whether the target teacher actually performs classroom observations, so they are not providing targeted feedback to the peer teachers on how to improve their teaching.

Additional Findings from This Evaluation

This sub-section shows results on outcomes that are not explicitly mapped to MoECRT's theory of change but are important indicators that could be impacted by the program.

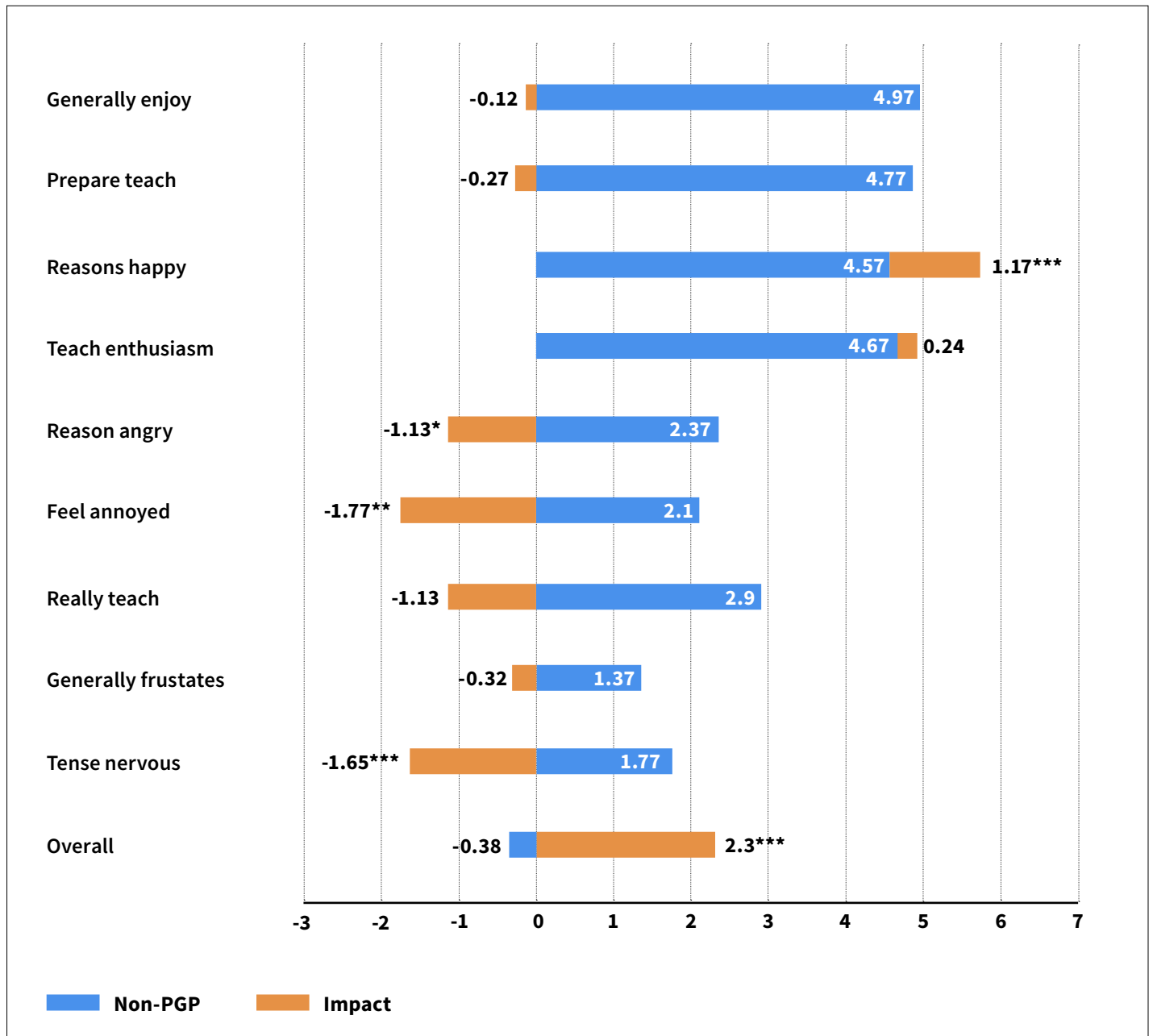
Figure 23 shows that the PGP teacher's self-reported enjoyment of teaching increased substantially. Teachers were asked to answer a battery of questions related to how much they enjoy teaching. The first row shows results for the index, and the subsequent row reports results for all the respective indicators, ranging from 1 "Highly disagree" to 5 "Highly agree". To construct the mean index, we first switch the values of the questions that were asked in a negative way, so that positive means more enjoyment of teaching. We then take the arithmetic mean of these indicators, standardized to a mean of zero and standard deviation of one. Teachers provide more positive responses to the statement "I often have

reasons to be happy while I teach". They are also less annoyed when teaching, and feel less tense, mad, and nervous when teaching. One possible interpretation is that teachers feel like they are more in control of their classroom, in line with the improvement in classroom culture.

This strong positive result on enjoyment of teaching is very encouraging from a sustainability point of view, since teachers who enjoy their job are more likely to persist in applying improved teaching practices.

Some questions could be asked of all teachers, including the 31 PGP teachers who had become school principals by the time of data collection. We could not conduct classroom observations for these teachers, because they are no longer teaching students. But we could ask about their career aspirations and self-confidence as a leader.

Table 12 report results using the full sample of targeted teachers, including the 31 PGP who had become school principals by the time of data collection. Recall that we could not conduct classroom observations with these teachers, because they are no longer teaching, but we asked all of them about their beliefs. Teachers who participated in the program are 44 percentage points more likely to become school principals, compared to those who do not. It also improved aspirations: teachers who participated in the program are 66 percentage points more likely to state that they want to become a school principal or inspector (or they had already become a school principal).

Figure 23. Enjoyment of teaching

Notes. See notes in Figure 15. Data come from the teacher survey. All dependent variables (except for the index) are Likert scale questions ranging from (1) "Strongly Disagree" to (5) "Strongly Agree". Overall is the arithmetic mean, standardized to have a mean of zero and standard deviation of one.

Table 12. Leadership confidence and aspirations

	Control Mean	Impact		
	(1)	(2) $ x < 3$	(3) $ x < 4$	(4) Full
Panel A. Leadership ambitions				
Became a school principal	0.06 (0.04)	0.06 (0.04)	0.14 (0.11)	0.02 (0.09)
Plan to be a school principal or inspector	0.19 (0.07)	0.19 (0.07)	0.84*** (0.26)	0.50*** (0.19)
Panel B, Self Confidence as a Leader				
Index	0.70 (0.23)	2.64** (1.07)	2.29*** (0.85)	1.61*** (0.57)
I have necessary ability to become a good leader of this school	3.68 (0.16)	1.26* (0.67)	1.38** (0.54)	0.66* (0.38)
I know what is needed to help teachers perform their duties	4.06 (0.11)	0.89 (0.56)	0.58 (0.43)	0.46 (0.30)
I believe in my ability to influence teachers	3.71 (0.14)	1.54** (0.67)	1.33** (0.54)	0.80** (0.36)
I know how to encourage teachers	3.84 (0.15)	1.52** (0.68)	1.22** (0.53)	0.70* (0.37)
Trainings have equipped me to become a school principal	3.10 (0.19)	1.63** (0.79)	1.57** (0.64)	1.55*** (0.49)
Observations	32	95	135	381

Notes. Data come from the teacher survey. The dependent variables in panel A are binary. All dependent variables (except for the index) in Panel B are Likert scale questions ranging from (1) "Strongly Disagree" to (5) "Strongly Agree". Index is the arithmetic mean of these indicators.

Their self-confidence as a leader improved substantially. These results are encouraging, since a large object of PGP is to train a new generation of school leaders.

Table 13 shows the same results as Table 12 but broken down by gender. Although statistical power is reduced with the sub-samples, there is suggestive evidence that program had a much larger impact on teachers' self-confidence for males, compared to females. The impact on a teacher's career aspirations also seems to be smaller for females relative to males.

Table 13. Leadership confidence and aspirations (by gender)

	Impact ($ x < 3$)		Impact ($ x < 4$)	
	(1) Male	(2) Female	(3) Male	(4) Female
Panel A. Leadership ambitions				
Became a school principal	3.23 (2.47)	0.36** (0.17)	-0.08 (0.36)	0.20 (0.12)
Plan to be a school principal or inspector	4.27 (3.53)	0.64** (0.27)	1.15* (0.67)	0.71*** (0.24)
Panel B. Self Confidence as a Leader				
Index—self-confidence as a leader	24.01 (30.41)	1.37** (0.68)	6.32** (2.71)	1.49** (0.65)
I have necessary ability to become a good leader of this school	6.27*** (1.75)	0.72 (0.63)	2.90* (1.56)	0.92* (0.50)
I know what is needed to help teachers perform their duties	16.23 (50.05)	0.64 (0.45)	1.98 (1.56)	0.48 (0.44)
I believe in my ability to influence teachers	35.12 (121.02)	0.75 (0.47)	5.49* (2.89)	0.73 (0.47)
I know how to encourage teachers	13.71 (18.11)	0.55 (0.40)	4.57** (1.98)	0.56 (0.38)
Trainings have equipped me to become a school principal	13.96 (19.33)	0.55 (0.80)	2.51 (1.56)	1.29* (0.72)
Observations	20	75	33	102

Notes. See table 10. Data come from the teacher survey. The dependent variables in panel A are binary. All dependent variables (except for the index) in Panel B are Likert scale questions ranging from (1) "Strongly Disagree" to (5) "Strongly Agree". Index is the arithmetic mean of these indicators.

However, this program did not have a positive impact on the current school principals' perceptions of the quality of the leadership ability of the PGP graduates who are still teachers. In fact, there is a large negative impact of 0.44 standard deviations on the leadership index. The largest reduction is in their belief in the ability of a teacher to influence and lead teachers. Note that the mean is extremely high, so the school principals almost always give everyone a value of five (out of five).

Table 14 School Principal's perception of the teacher's leadership ability

	Control Mean	Impact		
	(1)	(2) $ x < 3$	(3) $ x < 4$	(4) Full
Overall	4.77 (0.05)	-0.44 (0.27)	-0.55** (0.24)	-0.48*** (0.18)
Know how to be a good leader	4.80 (0.07)	-0.28 (0.40)	-0.43 (0.38)	-0.25 (0.25)
Helps team accomplish tasks	4.87 (0.06)	-0.43 (0.28)	0.69*** (0.29)	0.63*** (0.21)
Ability to influence and lead teacher	4.77 (0.08)	-0.70* (0.37)	-0.87*** (0.32)	-1.02*** (0.26)
Encourage good team	4.63 (0.10)	-0.34 (0.45)	-0.19 (0.35)	-0.11 (0.25)
Observations	31	89	127	350

7.



CONCLUSIONS AND RECOMMENDATIONS

The summarize, this report finds that the PGP program was well-implemented and positively received by teachers, whose teaching practices improved along a range of dimensions, including: application of positive discipline, classroom culture, and instruction. Teachers also increased their enjoyment of teaching. However, there is no evidence that the program improved teachers' socio-emotional competencies, nor that it increased differentiated instruction, as measured by MoECRT's own classroom observation instrument. Finally, there is also evidence that participants of the program engaged other teachers in the school in conversations around student-centered teaching and self-reflection. They are also more likely to become school principals, aspire to become school principals, and have more self-confidence in their leadership abilities.

The very large positive impacts on teaching practices, as measured using the Teach classroom observation instrument, are impressive. For comparison, recent study on teacher professional development in East Asia and Pacific found that programs implemented by government is often ineffective.³⁷ The government should be commended for designing and implementing a program that bucks this

trend. The fact that the impacts are different depending on the choice of classroom observation instrument merits further investigation. Both instruments cover very different domains of instructional quality and positive culture. The key difference might lie in the level of difficulty or complexity measured. MoECRT's instrument sets high standards for good teaching. These two instruments therefore measure very different points across the distribution of teaching practices.

Our findings present compelling evidence that the current program design and approach effectively supports teachers in enhancing their teaching practices, particularly in relation to their enjoyment in teaching and utilization of positive discipline with students. Despite ongoing efforts by the government to eradicate violence and discrimination in schools over the years, this program has successfully aided teachers in comprehending the principles of positive discipline and enhancing their classroom management skills. Through this program, teachers not only grasp the concept but also acquire the necessary guidance for implementing it in their classrooms. This achievement should be sustained.

³⁷ Afkar et al., 2023. The World Bank's regional flagship report collects information about various teacher professional development programs in East Asia and Pacific and found that these programs often lack the four element of effective teacher training program: (a) a focus on content knowledge, (b) opportunities to practice what is learned with colleagues, (c) continued support through follow-up visits focused on training content, and (d) career incentives through promotion or increased salary.

Moving forward, MoECRT should build on their successes, and bear closely in mind implementation challenges as the program gets scaled up in the next 18 months.

The results and impacts observed in the selected districts of Batch 5 (involving 8,000 participants from 166 districts) cannot be assumed to be replicated automatically if the government fails to maintain the quality of implementation in subsequent batches, such as Batch 8 which commenced with 11,730 participants from 284 districts in 34 provinces, as of May 2023.³⁸

Drawing upon the findings presented in this report, we provide the following recommendations to enhance the quality of PGP implementation.

A. Standardizing best practices

The primary focus of this recommendation is to urge policymakers to take the necessary actions to uphold quality of Guru Penggerak program (PGP) implementation across all districts. Our findings indicate that there is variation in teacher satisfaction, particularly concerning the logistical arrangements for workshop implementation. Despite the current provision of guidelines³⁹ by MoECRT to local governments and *Balai Guru Penggerak* on organizing PGP activities, there is still room for further improvement. For example, MoECRT should:

- Providing more detailed definitions of the "comfortable room" in the guidelines
- Implementing a checklist system to effectively monitor compliance with the intended standards.

Undoubtedly, successful implementation will rely on strong coordination between MoECRT and the local government/ Balai Guru Penggerak (BGP) operating in each district.

B. Enhancing Mentor's Preparedness

This improvement is crucial in ensuring that mentors are equipped with the necessary skills and knowledge to effectively support teachers in the short term.

Throughout the program, mentors play a vital role in facilitating teachers' comprehension of the training modules and providing assistance during workshops. However, their significance extends beyond these tasks as mentors serve as valuable coaches for teachers. Their responsibilities include conducting classroom observations, engaging in discussions with school principals regarding teachers' learning and providing feedback to enhance performance.

Moreover, it lays the foundation for a sustainable model of educational reform that can drive continuous improvement in the long term.

While the current recruitment process for mentors is open to the general public, MoECRT has planned to designate permanent mentors in each district from the pool of PGP graduates from previous batches. To ensure the effectiveness of mentorship, it is crucial to establish regular performance assessments and provide ongoing professional development opportunities for mentors. This approach will not only enable mentors to better support teachers participating in PGP in subsequent batches, but it also holds the potential to influence positive reform within their respective schools as they can implement and share the learned practices.

¹¹ This report finds that the PGP program was well-implemented and positively received by teachers, whose teaching practices improved along a range of dimensions, including: application of positive discipline, classroom culture, and instruction. However, there is no evidence that the program improved teachers' socio-emotional competencies, nor that it increased differentiated instruction..¹¹

³⁸ MoECRT, 2023

³⁹ Keputusan Direktur Jendral on GTK and Guidebook for facilitators, workshop, and individual coaching. Guideline for workshop available [here](#)

There are many ways to better support the PGP mentors, including but not limited to the following:

1. Provide Comprehensive Guidance for Mentors:

MoECRT has developed a comprehensive handbook for mentors, specifically focusing on their individual coaching responsibilities. While the handbook serves as a valuable resource, there is room for improvement, particularly in the utilization of classroom observation tools. These tools are vital for mentors to objectively observe teachers in the classroom and provide constructive feedback to enhance their teaching practices. However, the current classroom observation tools used in the program lack detailed guidance on how to use it and the scoring. By incorporating this detailed guidance, mentors will not only be better equipped to utilize the tools effectively within the program but also extend their application to their respective roles within their schools, offering coaching to other teachers. This enhancement will foster improved teacher support and development, contributing to the overall success of the program.

2. Enhance Mentor Training: MoECRT should invest in enhancing the training provided to mentors. This should include incorporating additional pedagogical strategies, leadership skills, and effective mentoring techniques. Additionally, teachers should also receive assistance in utilizing the tools included in the handbook. By equipping mentors with a broader range of knowledge and skills, they will be better prepared to support teachers in their professional growth.

3. Ensure Sustainable Mentor Support: MoECRT should establish a robust support system for mentors, including regular mentoring meetings, peer learning opportunities, and access to resources. This support network will help mentors overcome challenges, share experiences, and continuously enhance their effectiveness in guiding teachers.

By implementing these measures, it is expected that MoECRT can significantly improve the preparation of PGP mentors, enabling them to perform their tasks more effectively in the short term. Furthermore, this investment in mentorship will serve as a sustainable model for driving continuous reform and improvement within the education system in the long run.

C. Supporting Teachers in Differentiated Learning and Socio-Emotional Skills

To further enhance the program's effectiveness, it is recommended to provide teachers with additional support in differentiated learning and socio-emotional skills. This can be achieved through additional development of guidelines and tools, such as examples tools for diagnostic assessments, lesson plans with practical tips and suggestions for how to introduce differentiated learning in the classroom. Support can also encompass additional mentoring sessions during the program and follow-up mentoring sessions after the program ends.

D. 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. In fact, a review of the evidence showed that programs targeting socio-emotional skills of adults have limited to no impact. Moreover, the official standards for reaching competence in the different domains, as defined in the classroom observations instrument, are highly ambitious. For example, it might be counter-productive to expect the following:

“Teacher displays a consistently positive expression of emotions (displays happy, enthusiastic, friendly emotions)”

“Teachers smoothly modify strategies, materials, and groupings to optimize students' opportunities to learn and meet their learning need”

“Students are actively involved throughout learning and focused on work that not only develops skills and challenges them to be problem solvers.”

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, including:

- No longer including a goal of improving teachers' socio-emotional competencies.
- Reduce the definition of what entails differentiated instruction, focusing on conducting student assessments and adapting the level of teaching to students' levels of ability, rather than also adapting the teaching style to the learning needs of the child. The latter is extremely difficult for teachers to implement.
- Changing the targets for excellent teaching in the classroom observation instrument.

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.

The "Teachers' Learning Journey"⁴⁰ developed by MoECRT serves as an opportunity to critically assess whether the ideal teacher envisioned is grounded in rationale and practicality. This evaluation will enable the identification of areas that require supplemental interventions to support sustained and lasting changes in teachers' professional development.

E. Further research

As discussed in the limitations section above, the results are not generalizable to Indonesia as a whole, since the sample only includes primary school teachers in five districts in Java that were short-listed for Batch 5 training. This was done for cost considerations. Moreover, despite the relatively large sample of 350 teachers, statistical power is limited due to some non-compliance and the RDD approach which prioritizes results from teachers very close to the threshold for selection. We therefore do not have enough statistical power to perform sub-group analysis. There are also no results on student-level outcomes. Additional data collection and further our understanding of the broader impact of the program.

1. Conduct a randomized control trial. A randomized controlled trial will provide stronger evidence of the causal impact of the program, strengthen the external validity of the study, and also increase statistical power relative to a RDD. This could be implemented through a randomized phased in approach, where teachers assigned to the control participate in a later batch.

2. Expand data collection to more districts, including more remote areas. Since implementation is decentralized, it is possible that the impacts of the program will vary substantially by region.

3. Expand data collection to teachers who participated in earlier and later batches of the program. The program is scaling up over time—from 2800 teachers who participated in Batch 1 to 55,000 teachers who were selected in Batch 10—and there is international evidence that the effectiveness of programs decreases with scale. The characteristics of the teachers selected in the different batches might also be different, since it was more selective at the start. All of this means that the program might have had a larger impact in the earlier batches, relative to later batches.

4. Measure student-level cognitive and non-cognitive skills. An important next step is to ascertain whether the improvements in teaching practices translated into improvements in student outcomes. Given the improvements in application of positive discipline in the classroom, it is possible that the program improved both cognitive skills and other non-cognitive skills such as social behavior and self-confidence.

5. Longitudinal Study: Implement a longitudinal study design that captures data before, immediately after program completion, and tracks teachers' evolution over time, particularly as they transition into school principal roles. Additionally, utilizing national assessment data on the learning environment can offer valuable insights.

By undertaking future research with these considerations, a more comprehensive and nuanced understanding of the program's impact on teachers and the education system as a whole can be achieved.

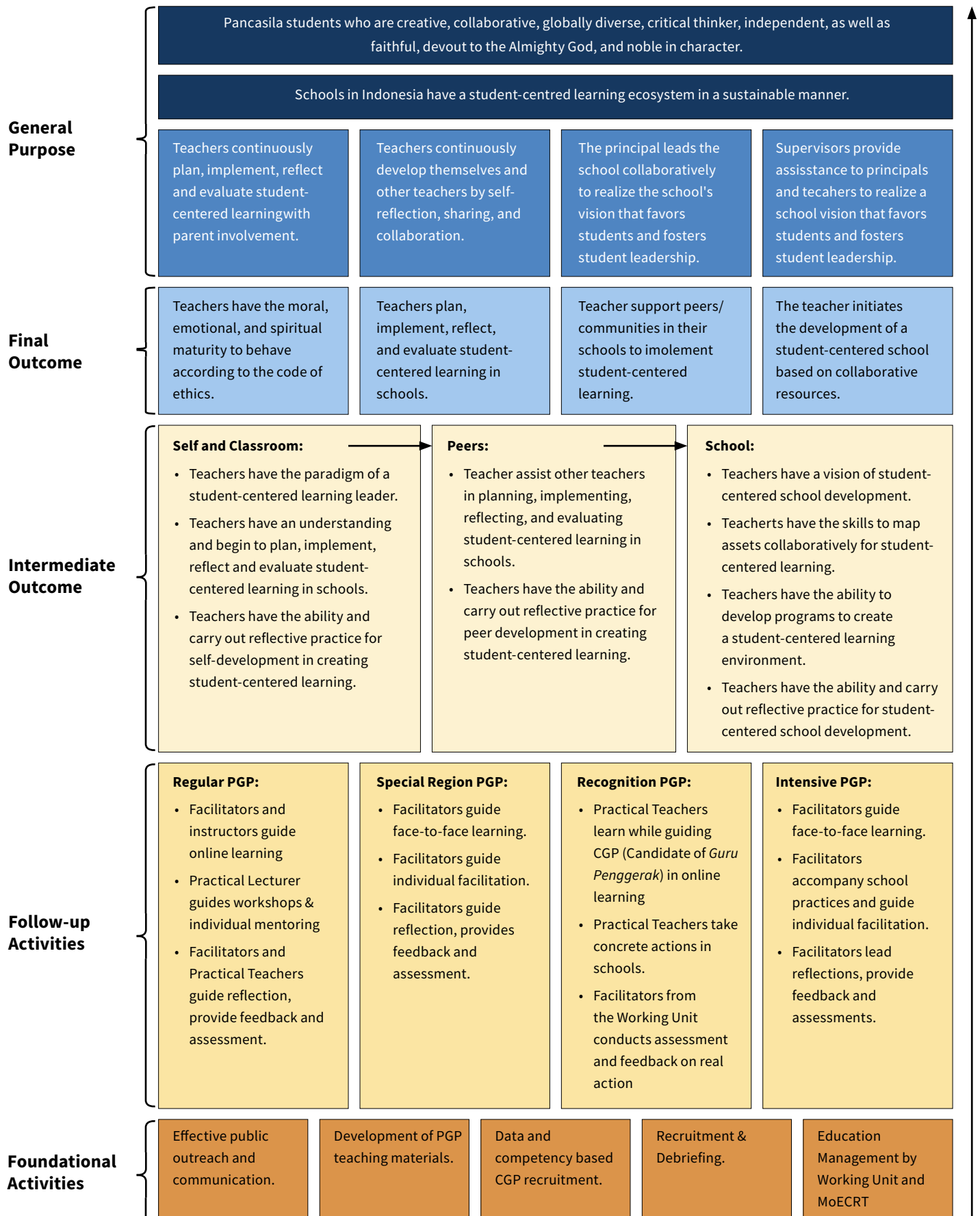
⁴⁰ Official definition is currently no available. MoECRT often refers 'teacher learning journey' as part of teacher's development following the paradigm of 'Merdeka Curriculum', which means a comprehensive self-reported assessment of individual's teacher competence and their improvement plan, for example here: <https://gurudikdas.kemdikbud.go.id/news/modul-pelatihan-peningkatan-kompetensi-numerasi-untuk-guru>

REFERENCES

- Afkar, R., Béteille, T., Breeding, M., Linden, T., Mason, A., Mattoo, A., Pfitze, T., Sondergaard, L., and Yarrow, N. 2023. Fixing the Foundation: Teachers and Basic Education in East Asia and Pacific. *World Bank East Asia and Pacific Regional Report*. Washington, DC: World Bank. doi:10.1596/978-1-4648-1904-9.
- Baeten, M., Kyndt, E., Struyven, K., Dochy, F. 2010. Using student-centered learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, 5(3), p.243-260
- GTK. 2023. Sapa GTK Episode 12: Guru Penggerak, as cited in *Kemendikbudristek Dorong Guru Penggerak Pimpin Transformasi di Satuan Pendidikan*. <https://gtk.kemdikbud.go.id/read-news/kemendikbudristek-dorong-guru-penggerak-pimpin-transformasi-di-satuan-pendidikan>
- Katadata. 2022. *Mayoritas Kepala Sekolah SD Berumur Lebih dari 55 Tahun*. <https://databoks.katadata.co.id/>
- Kim, J. 2004. Education reform policies and classroom teaching in South Korea, *International Studies in Sociology of Education*, 14(2), p.125-146, doi: 10.1080/09620210400200122
- Lee, S. Chia, A., Pereira, A., Tay, L. 2022. Singapore's Student-Centred, Values-Driven education system: A case study of Teacher Professional Learning. *Centering Whole-Child Development in Global Education Reform* (1), p.21. Routledge
- Ministry of Religious Affairs (MoRA). 2023. EMIS Kemenag 2022/2023. <https://emis.kemenag.go.id>
- MoECRT. 2020. *Merdeka Belajar Episode 5: Guru Penggerak*. <https://www.kemdikbud.go.id/main/blog/2020/07/kemendikbud-luncurkan-merdeka-belajar-episode-5-guru-penggerak>
- MoECRT. 2023. *Kurikulum Merdeka*. <https://kurikulum.kemdikbud.go.id/kurikulum-merdeka/>
- MoECRT. 2023. *Sebanyak 11.730 Guru Ikuti Pendidikan Guru Penggerak Angkatan 8*. <https://www.kemdikbud.go.id/main/blog/2023/05/sebanyak-11730-guru-ikuti-pendidikan-guru-penggerak-angkatan-8>
- MoECRT. 2023. Unduhan. <https://sekolah.penggerak.kemdikbud.go.id/gurupenggerak/unduh>
- MoECRT Education Statistics. 2023. *Statistik Pendidikan*. <https://statistik.data.kemdikbud.go.id>
- OECD. 2005. *Teachers Matter*. <https://www.oecd.org/education/school/34990905.pdf>
- OECD. 2018. *Programme for international student assessment (PISA) 2018*. https://www.oecd.org/pisa/publications/PISA2018_CN_IDN.pdf
- OECD. 2020. *Teaching and Learning International Survey (TALIS) 2018 Insights and Interpretations*. https://www.oecd.org/education/talis/TALIS2018_insights_and_interpretations.pdf
- Park, R. 2016. *Preparing students for South Korea's creative economy: The success and challenges of educational reform*. Vancouver B.C.: Asia Pacific Foundation of Canada.
- Permendikbud No.40. 2021. *Peraturan Menteri Nomor 40 tentang Penugasan Guru sebagai Kepala Sekolah*. https://jdih.kemdikbud.go.id/detail_peraturan?main=2940
- Permendikbud No.26. 2022. <https://peraturan.bpk.go.id/Home/Details/224728/permendikbudriset-no-26-tahun-2022>
- Puerta, M., Valerio, A., and Bernal, G. 2016. *Taking stock of programs to develop socioemotional skills: A systematic review of program evidence*. doi: 10.1596/978-1-4648-0872-2
- Schweisfurth, M. 2013. *Learner-centred Education in International Perspective Whose pedagogy for whose development?*. New York: Routledge
- World Bank. 2020. *The Human Capital Index 2020 Update: Human Capital in the Time of COVID-19*. © World Bank, Washington, DC.
- World Bank. 2023. *The Invisible Toll of COVID-19 on Learning. Indonesia Economic Prospects*. © World Bank, Washington, DC.
- Yang, J. and Tan, C. 2019. Advancing student-centric education in Korea: Issues and challenges. *The Asia-Pacific Education Researcher*, DOI: 10.1007/s40299-019-00449-1, 1-11.
- Yarrow, N., Afkar, R., Masood, E., Gauthier, B. 2020. *Measuring the Quality of MoRA's Education Services (English)*. Washington, D.C. : World Bank Group.

APPENDIX 1 – PENDIDIKAN GURU PENGGERAK

A. Theory of Change



B. Guru Penggerak program: Curriculum

Curriculum overview

Guru Penggerak Program Curriculum		
1 Paradigm and Vision of Guru Penggerak 1. Reflection on the National Philosophy of Education 2. Values and Roles of <i>Guru Penggerak</i> 3. Vision of <i>Guru Penggerak</i> 4. Positive Culture	2 Student-centered Learning Practices 1. Meeting Student's Learning Needs 2. Social and Emotional Learning 3. Coaching for Academic Supervision	3 Leadership in School Development 1. Value-based Decision Making 2. Leadership in Resource Management 3. Management of Impactful Programs for students
Workshop: 0. Orientation 1. Community Practitioner Development 2. Vision for Environmental Learning Change	Workshop: 3. Leader's Role in Learning Development 4. Strengthening Coaching Practices	Workshop: 5. Collaboration in Student-centered Management 6. Sustainability of School Development 7. Final Learning Outcomes
Focus on self and classroom improvement of <i>Guru Penggerak</i> Candidate	Focus on peer teachers improvement of <i>Guru Penggerak</i> Candidate	Focus on school improvement of <i>Guru Penggerak</i> Candidate

Final outcomes and derivative indicators, as per MoECRT's theory of change.

Major grouping	Grouping	Derivative Indicators	
I. Teachers have the moral, emotional, and spiritual maturity to behave according to the code of ethics.	Teacher has the skills and habits of reflection, then uses them as the basis for the preparation of a follow-up plan that reflects partiality to the students	I.1a	Teacher can periodically interpret facts/events related to the learning experience in the development of self-competence by describing the feelings involved, describing the learning, and making improvement/decision plans to create a student-centered learning environment.
		I.1b	Teacher can periodically interpret facts/events related to the teaching experience with students by describing the feelings involved, describing learning, and making improvement/decision plans to create a student-centered learning environment.
	Teacher analyzes their position in leadership competencies and have a self-development plan according to their individual needs	I.2a	Teacher is able to study and infer their own position in the competency map of the <i>Guru Penggerak</i>
		I.2b	Teacher has documented plans and targets for self-development based on the identification of needs for increased competence
		I.2c	Teacher carries out the prepared self-development plans and targets

Major grouping	Grouping	Derivative Indicators	
	Teacher masters social-emotional competence in teaching practice in schools, both in the intracurricular, co-curricular and extracurricular realms.	I.3a	Teacher is able to identify the emotions felt in an event.
		I.3b	Teacher responds to emotions in a positive, healthy, and productive way; build in the achievement of goals in the learning process of students
		I.3c	Teacher shows empathy regarding the student's point of view
		I.3d	Teacher shows efforts to be a resilient in dealing with challenging situations in social relations in the school environment
		I.3e	Teacher carries out the decision-making process with empathy, is aware of the consequences, and is student-centered.
	Teacher practices decision-making based on virtue values as a leader who favors students	1.4a	Teacher applies decision-making and testing measures in favor of the student in situations of ethical dilemmas using moral persuasion.
II. Teachers plan, execute, reflect on, and evaluate student-centered learning in schools.	The teacher applies positive discipline in the classroom and in school.	II.1a	Teacher engages students in crafting class beliefs
		II.1b	Teachers apply monitors or managers positions consistently in the learning process in the classroom and school.
		II.1c	Teacher applies the restitution process to the students when making a mistake.
	Teachers apply differentiation strategies in the learning process in the classroom and school (preparation; teaching, to assessment).	II.2a	Teacher able to identify and meet students' learning needs based on data.
		II.2b	Teacher able to apply content differentiation strategies, process differentiation, and product differentiation in learning (preparation, teaching and assessment)
	Teachers practice learning that is integrated with social-emotional competence.	II.3a	Teacher structures learning to improve students' social emotional skills.
		II.3b	Teacher able to use various strategies in integrating social and emotional learning.
III. Teacher encourages peers/community to implement student-centered learning.	Teacher develops learning communities with colleagues in their schools to increase competence to carry out student-centered learning.	III.1a	Teacher establishes a learning community with colleagues at his school.
		III.1b	Teacher and their peers have a map of needs within the learning community at school to enhance student-centered learning.
		III.1c	Teachers and their peers have self-development plans based on a map of needs.
		III.1d	Teacher and their peers run learning communities according to development plans.

Major grouping	Grouping	Derivative Indicators	
	Teacher applies coaching principles in empowering their peers to create student-centered learning.	III.2a	Teacher works with colleagues to determine goals in the coaching process in order to support student-centered learning.
		III.2b	Teacher assists peers to identify student-centered development needs
		III.2c	Teacher guides their peers in creating action plans for student-centered learning development
		III.2b	Teacher uses coaching skills in carrying out academic supervision
IV. Teacher initiates student-centered school development with collaborative resource (asset) based.	Teacher has a vision of student-centered school development.	IV.1a	Teacher involves various actors to reflect on the school's vision to be more partial to students
		IV.1b	Teacher, colleagues and principal have a vision of student-centered school development based on the results of joint reflection
	Teacher initiates resource-based student-centered school development, involving school residents.	IV.2a	Teacher strives for a school environment that fosters student leadership (voice, choice, and ownership) in the development of school programs.
		IV.2b	Teacher provides opportunities for students to demonstrate voice, choice, and ownership in the development of school programs
		IV.2c	Teacher able to identify and use a variety of assets (7 main capitals) according to the school context to initiate sustainable and student-centered school development.
		IV.2d	Teacher engages a variety of actors in schools with clear roles and responsibilities in structuring and carrying out the initiation of sustainable and student-centered school development programs.

C. Applicants and Selected Candidates, Batch 1 to 5

Participants	Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
Districts	56	56	56	166	166
Applicants	19,218	17,091	23,274	42,009	105,643
Verification and validation of Application	5,142	6,401	6,662	16,366	49,815
Stage 1 Selection: <i>Essay</i>	4,598	5,217	4,446	10,826	19,974
Stage 2 Selection: <i>Interview and Teaching simulation</i>	2,800	2,800	2,801	8,053	8,000 (expected)
Acceptance rate	15%	16%	12%	19%	8%
Completed the program	2395	3,004	2760	7948	7391

D. PGP Supporting Actors⁴¹

	Mentor	Facilitator	Instructor
Participants	Teachers, PGP Alumni, or school principals with minimum of 5 years teaching experience and 1 year of mentoring experience	Civil servant trainer or school supervisor with minimum of 5 years teaching experience	Former PGP facilitators with minimum of 5 years teaching experience
Roles	1. Become mentor for CGP to: a) Facilitate workshop b) Visit school c) Assist in making follow-up plans 2. Coordinate with working group 3. Coordinate with facilitator 4. Collect and analyze CGP Principal and peer's feedback and communicate to CGP 5. Report CGP's achievements	1. Facilitate the online learning activities: a) Learning reflection b) Concept exploration c) Collaboration Room 2. Provide feedback for all CGP's assignments 3. Coordinate with working group 4. Coordinate with mentor 5. Make work report	1. Provide learning materials to CGP 2. Sharing good practice of applying the material contained in PGP modules 3. Become a resource person for Mentor and Facilitator

Source. PGP Batch 5 Master Deck, Permendikbud No.26 (2022) on Pendidikan Guru Penggerak, and Sekolah Penggerak Website: PGP Actors Selection Process

Notes. for the upcoming batches MoECRT plan to prioritize PGP alumni to become mentor and facilitator.

E. PGP Classroom Observation Instrument

1. Opening of the Learning Process	1. The teacher immediately starts the learning without mentioning the purpose of the learning. 2. The teacher only conveys the learning objectives and does not explain the activities. 3. The teacher ensures that students understand the expected learning objectives and the activities. 4. The teacher ensures that students understand the expected learning objectives and the activities and discusses with students the relationship of topics discussed with other materials and with daily life.
2. Comfortable Learning Environment to Meet Student Learning Needs	1. Teacher does not show any modification of the learning environment according to learning objectives, characteristics of students and social emotional competence. 2. Teacher modifies the learning environment according to learning objectives (academic learning needs). 3. Teacher modifies the learning environment according to learning objectives and understanding student characteristics. 4. Teacher modifies the learning environment according to learning objectives, student characteristics, and social and emotional learning objectives.
3. Positive Communication Usage: Interaction	1. Teacher dominates the interaction and gives few opportunities for students to express an opinion (one-way lecture method). 2. During one class hour, half of teacher communication is one-way and half of the time students have the opportunity to express an opinion.

⁴¹ PGP Masterdeck PPT, Sekolah Penggerak [Website](#)

	- Teacher communicates reciprocally with student (teacher-student). - Teacher communicates reciprocally with students (teacher-student) and gives students the opportunity to interact with each other (students).
4. Positive Communication Usage: Empathy	- The teacher makes assumptions about the condition of the pupil without clarifying first. - The teacher listens to the opinions of students that are relevant to the topic or questions asked by the teacher alone. - The teacher tries to figure out what their students are working on and thinking, but only manages to do it with a few students. - The teacher finds out what their student is working on and thinking. The teacher listens actively to the opinions expressed by all students.
5. Positive Communication Usage: Positive emotions	- Teacher uses negatively charged words. - Teacher displays inconsistent expressions of emotions (between negative and positive emotions). - Teacher displays positive expressions of emotions and neutral emotions alternately. - Teacher displays a consistently positive expression of emotions (happy, enthusiastic, friendly emotions).
6. Application of Discipline	- The Punisher Position: Teacher uses punishment, innuendo, and criticism to discipline students. The teacher will express anger when the student makes a mistake by rebuking, yelling, or pointing. - Position which makes students feel guilty: The teacher silences the student or uses words that make the student feel guilty (as if the teacher/parent/other students will suffer as a result of their mistake). - Friend Position: Teacher uses friendship and humor to influence students. Teacher defends, justifies or provides explanations for student behavior. - Monitor Position: Teacher applies discipline by using the rules and consequences agreed upon with the pupil, using objective evidence. The consequences given relate to the rules that the student violated. - Manager Position: Teacher uses the principle of restitution that reasonable mistakes occur, and they accept the mistakes of students. The teacher asks the students why they made a mistake and help them find a way out. This type of teacher sometimes uses the position of monitor and manager in students in certain situations.
7. Motivation	- Teacher pushes students to learn by providing negative punishments/consequences. - Teacher encourages student learning behavior by conveying teacher expectations and providing positive consequences. - Teacher encourages students by reminding them of the importance of learning. - Teacher fosters students' desire to learn by providing engaging and interesting learning sessions.
8. Socio-Emotional Content in Learning Activities	- The learning activities provided are only academically charged. - There are activities that make students excited, having fun, or refocus on learning (such as: ice breaking or energizers) but are not explicitly utilized to strengthen social-emotional competence and are separate from the academic content that is being learned. - Activities that reinforce students' socio-emotional competence are integrated into learning at: a warm opening stage, a core involving-challenging-fun stage, and an optimistic but still separate closing stage from the academic material that is being learned.

	4. Activities that reinforce students' socio-emotional competence have been integrated into learning at: a warm opening stage, a core involving-challenging-fun stage, and an optimistic closing stage by fusing it with academic content being studied.
9. Approaches to Accompanying Students	1. Teacher is silent and allows the students to do the assigned tasks themselves. 2. Teacher provides direction and solutions directly to students when they work on assignments. 3. Teacher helps students who are experiencing difficulties. 4. Teacher helps students according to their learning needs. Students can do their assignments independently as a result of teacher support.
10. Use of Learning Strategies	1. Teacher uses only one learning strategy throughout one session. OR Teacher uses learning strategies but find it difficult to actively engage students in learning. 2. Teacher uses several learning strategies and tries to encourage students to be actively involved in learning and encourage skills development , but not consistently. 3. Teacher uses several relevant learning strategies to actively engage students throughout learning and encourage the development of important skills. 4. Teacher smoothly modifies strategies, materials, and groupings to optimize students' opportunities to learn and meet their learning needs. Students are actively involved throughout learning and focused on work that not only develops skills but challenges them to solve problems.
11. Closing of the Learning Process	1. Teacher closes the class without giving learning conclusions. 2. Teacher only focuses on mentioning the teaching materials at the session meeting but does not attempt to conclude other learnings that were non-academic. 3. Teacher concludes and tells the student what other learning is gained besides the teaching material (e.g., related to the social emotional skills learned). 4. Teacher invites students to conclude, reflect on what has been learned (both academically, socially, and emotionally), and appreciate the positive progress or changes made by students.
12. Providing Feedback	1. Teacher gives feedback that may offend the student personally (example: blaming the student for their nature). 2. Teacher only gives general feedback on the product of the whole student. OR The teacher only gave a brief comment. 3. Teacher gives clear, concrete feedback) to the student's product that fits the learning objectives. 4. Teacher provides constructive feedback (clear and containing development advice) not only on the student's product but also their learning process (e.g., how the student makes thinking strategy, how the pupil solves the problem, communication techniques that was done by students).
13. The Functions of the Assessment	1. During the learning process, there is no assessment process. 2. During the learning process, there is an assessment process, but it is not closely related to the learning objectives. 3. During learning, an assessment process occurs and there is evidence of assessment according with the learning objectives. 4. As learning progresses, an assessment process occurs that is according with the learning objectives. There is evidence of teachers' efforts to differentiate to meet students' learning needs and ensure students understand the assessment process.

APPENDIX 2 – TEACH CLASSROOM OBSERVATION INSTRUMENT:

C. Applicants and Selected Candidates, Batch 1 to 5

Area	Element (Score 1-5)	Behavior (Score Low, Medium, High)
Time on Task	0. Time on Learning	0.1 Teacher provide activities to most students 0.2 Students are on task
A. Classroom culture	1. Supportive Learning Environment	1.1 Teacher treats all students respectfully 1.2 Teacher uses positive language with students 1.3 Teacher responds to students' needs 1.4 Teacher does not exhibit bias and challenges stereotypes in the classroom
	2. Positive Behavioral Expectations	2.1 Teacher sets clear behavioral expectations for classroom activities 2.2 Teacher acknowledges positive student behavior 2.3 Teacher redirects misbehavior and focuses on the expected behavior, rather than the undesired behavior
B. Instruction	3. Lesson Facilitation	3.1 Teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives 3.2 Teacher explains content using multiple forms of representation 3.3 Teacher makes connections in the lesson that related to other content knowledge or students' daily lives 3.4 Teacher models by enacting or thinking aloud
	4. Checks for Understanding	4.1 Teacher uses questions, prompts or other strategies to determine students' level of understanding 4.2 Teacher monitors most students during independent/group work 4.3 Teacher adjusts teaching to the level of students
	5. Feedback	5.1 Teacher provides specific comments or prompts that help clarify students' misunderstandings 5.2 Teacher provides specific comments or prompts that help identify students' successes
	6. Critical thinking	6.1. Teacher asks open-ended questions 6.2. Teacher provides thinking tasks 6.3. Students ask open-ended questions or perform thinking tasks
C. Socio-emotional skills	7. Autonomy	7.1. Teacher provides students with choices 7.2. Teacher provides students with opportunities to take on roles in the classroom 7.3. Students volunteer to participate in the classroom
	8. Perseverance	8.1. Teacher acknowledges students' efforts 8.2. Teacher has a positive attitude towards students' challenges 8.3. Teacher encourages goal setting
	9. Social and collaborative skills	9.1. Teacher promotes students' collaboration through peer interaction 9.2. Teacher promotes students' interpersonal skills 9.3. Students collaborate with one another through peer interaction

TEACH Overall Result

This section will explain the teaching practices score of the entire sample using the Teach Primary observation tool. The results will combine both the treated and control groups of our sample, as the aim is to assess the overall result rather than making comparisons. It is important to note that the sample in this research are limited to teachers who were applying for participating in Guru Penggerak program – Batch 5, and hence most likely consists of high performers and highly motivated teachers in Java, the most populated island in Indonesia. Therefore, the Teach Primary score below does not represent the teaching practices of Indonesian teachers because the sample is not nationally representative.

Figure 24. TEACH Result - Time on learning

Time on Learning	Teachers provide learning activity	No	Yes		
		5%	95%		
	Students are on task	Low		Medium	High
		6 or more students are off task	2 to 5 students are off task	0 or 1 students are off task	
		7%	23%	65%	

Teachers in our sample provide learning activities to students 95% of the time. They engage students with various activities that related to learning such as providing lecture or worksheet, facilitating individual and group work. However, when teachers provide task, the entire class are on task only 65% of the time. The rest of the time, at least two students did not participate by being distracted or causing disruption in the classroom.

Data collected using Teach Tool indicates that 1 out of every 10 teachers struggle (score less than 2) in at least one area. Results from Figure 25 above indicates that teachers have strong ability in Classroom Culture (94% of teachers have a score higher than 3) and in Instruction (64% of teachers scored higher than 3); however, they exhibit slightly weaker ability in Socioemotional Skills (58% of teachers have a score less than 3).

Overall, teachers performed well in the Classroom Culture area, achieving an average score of 3.6 out of 5 possible points. They demonstrated moderate performance in the Instruction area, with an average score of 3.1, and weaker performance in the Socioemotional Skills area, with an average score of 2.7.

Figure 25. Teach Primary Score of All Areas

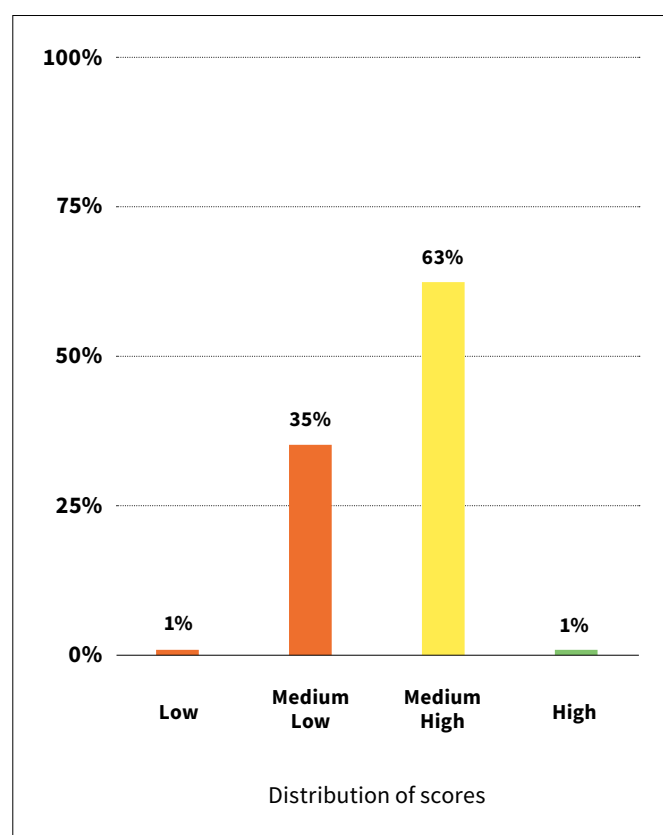


Table 15. TEACH Results

Variable	(1) Total Mean/(SE)/SD	(2) Female Mean/(SE)/SD	(3) Male Mean/(SE)/SD	(2)-(3) Pairwise t-test Mean difference
Classroom Culture	3.569 (0.019) 0.507	3.565 (0.022) 0.490	3.578 (0.038) 0.547	-0.013
Supportive Learning Environment	3.660 (0.022) 0.572	3.672 (0.026) 0.568	3.631 (0.041) 0.584	0.041
Positive Behavioral Expectations	3.477 (0.029) 0.765	3.457 (0.033) 0.741	3.524 (0.057) 0.819	-0.067
Instruction	3.103 (0.022) 0.579	3.111 (0.026) 0.579	3.085 (0.040) 0.581	0.026
Lesson Facilitation	3.713 (0.029) 0.767	3.719 (0.036) 0.798	3.699 (0.048) 0.689	0.020
Checks for Understanding	3.347 (0.034) 0.911	3.356 (0.041) 0.918	3.325 (0.063) 0.898	0.031
Feedback	2.186 (0.039) 1.041	2.164 (0.047) 1.039	2.238 (0.073) 1.048	-0.074
Critical Thinking	3.167 (0.037) 0.990	3.204 (0.043) 0.964	3.078 (0.073) 1.047	0.127
Socioemotional Skills	2.678 (0.026) 0.679	2.677 (0.031) 0.697	2.678 (0.044) 0.634	-0.001
Autonomy	2.941 (0.036) 0.944	2.968 (0.043) 0.953	2.879 (0.064) 0.921	0.089
Perseverance	2.283 (0.023) 0.600	2.277 (0.028) 0.613	2.296 (0.040) 0.571	-0.019
Social and Collaborative Skills	2.809 (0.051) 1.336	2.787 (0.062) 1.372	2.859 (0.087) 1.247	-0.072
TEACH Global Score	3.116 (0.017) 0.438	3.118 (0.020) 0.436	3.114 (0.031) 0.441	0.004
Number of observations	700	494	206	700

Significance: ***=.01, **=.05, *=.1.

Table 16. Overview of TEACH Instrument

Number of Schools	285
% of PGP school	10%
% of Non-PGP school	90%
% of Urban school	25%
% of Rural school	75%
Number of Teachers	350
# of TEACH Primary Observations	700

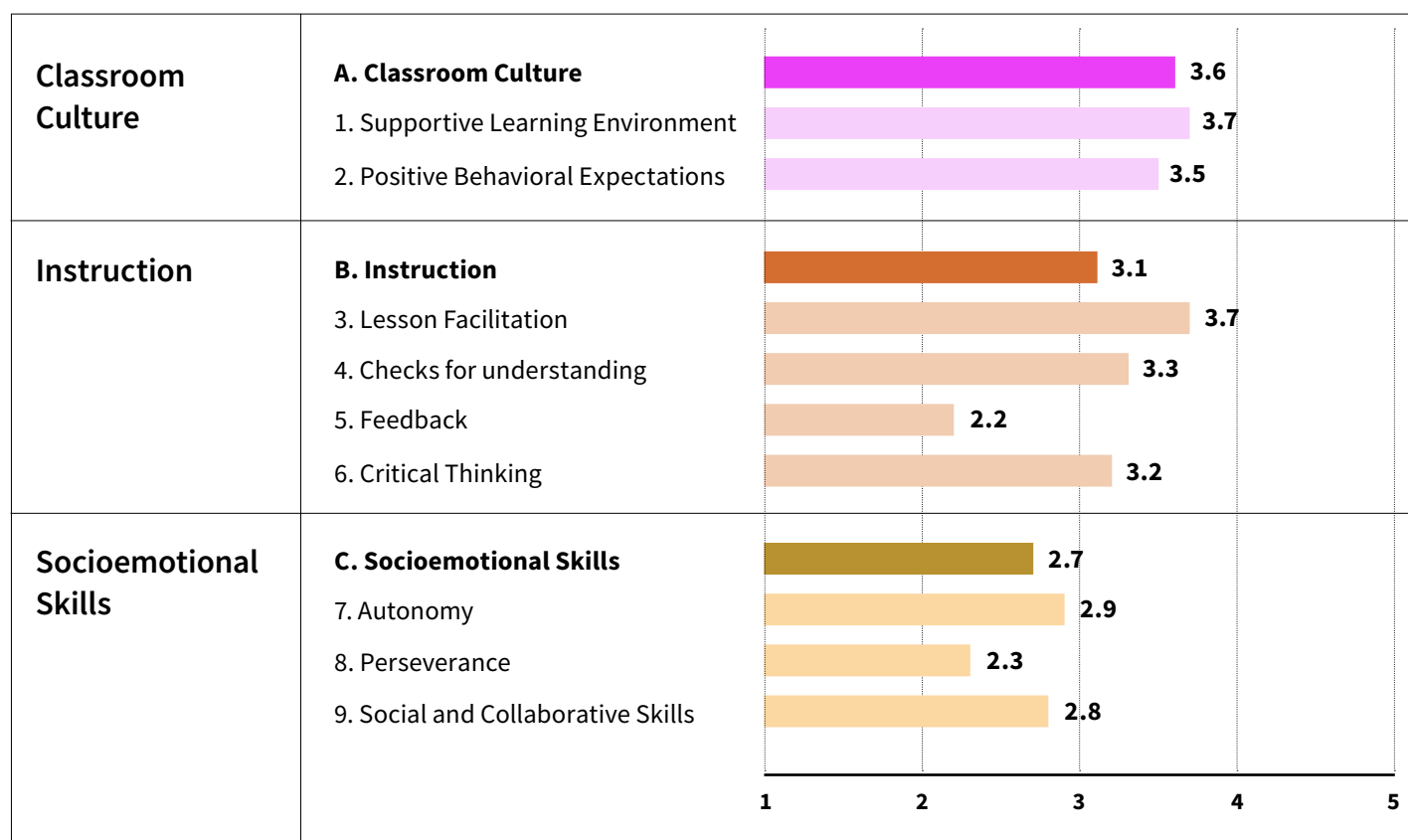
Figure 26. Distribution of Average Teach Primary Scores by Area and Element

Figure 26 shows that Indonesian teachers in our sample are relatively skilled at creating a supportive learning environment and setting positive behavioral expectations. Moreover, these teachers are also skilled in facilitating the lesson. They score around the medium range in checking for understanding and encouraging students to think critically; but are less skilled at providing feedback. Lastly, teachers are around the medium range in promoting student autonomy and social and collaborative skills; but are poor at fostering perseverance.

Overall, teachers performed well in the Classroom Culture area, achieving an average score of 3.6 out of 5 possible points. They demonstrated moderate performance in the Instruction area, with an average score of 3.1, and weaker performance in the Socioemotional Skills area, with an average score of 2.7.

APPENDIX 3 – EMPIRICAL STRATEGY

We estimate the Local Average Treatment effect using a two-stage least squared approach. The first and second stage equations are respectively:

$$PGP_{i,d} = \alpha_0 + \sum_{p=1}^2 \alpha_{1,p} x_{i,d}^p + \alpha_2 Above_{i,d} + \sum_{p=1}^2 \alpha_{1,p} (x \times Above)_{i,d}^p + X_i + \gamma_d + \epsilon_{i,d} \quad (1)$$

$$y_{i,d} = \beta_0 + \sum_{p=1}^2 \beta_{1,p} x_{i,d}^p + \beta_2 PGP_{i,d} + \sum_{p=1}^2 \beta_{1,p} (x \times Above)_{i,d}^p + X_i + \gamma_d + \epsilon_{i,d} \quad (2)$$

Where $y_{i,d}$ is the outcome of interest for teacher i in district d ; $x_{i,d}$ is the running variable—teachers' score in the selection tests, subtracted by the cut-off value for their district and contract type; and $Above_{i,d}$ a dummy variable equal to one if a teacher crosses threshold for selection in their district; $PGP_{i,d}$ is a dummy variable equal to one if a teacher was selected to participate. X is a vector of teacher-level pre-treatment or time-invariant characteristics, such as age, teacher contract type, teachers' education level, and gender.

Our estimate of the local average treatment effect is β_2 . Since we have one-sided non-compliance, with a zero probability of teachers being selected if they fall just short of the threshold, this estimate can also be interpreted as the Treatment Effect on the Treated: the impact of being selected to participate in PGP.

For robustness, we show different estimates, varying the size of the bandwidth and the level of polynomial relationship between the running variable and the dependent variable. In particular, we include a quadratic fit when estimating the model using the full sample of teachers, but a linear fit when we restrict the bandwidth. We show estimates for bandwidths that range from just two values from the cutoff to six values from the cutoff.

Control variables are selected using the adapted Lasso procedure. We select control variables that are both predictive of treatment and the dependent variable of interest.

The F-statistic for the first-stage regression is very strong: 146.54.

APPENDIX 4 – ADDITIONAL TABLES AND RESULTS

Table A.4.1: Teacher Characteristics (incl. recent school principals)—by PGP

Variable	(1) PGP Mean/SE	(2) Control Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Administrative data				
Selection score	100.152 (0.240)	91.385 (0.309)	96.990 (0.288)	8.794***
Female	0.697 (0.029)	0.796 (0.035)	0.732 (0.024)	-0.099**
Civil servant teacher	0.717 (0.029)	0.533 (0.043)	0.651 (0.024)	0.184***
Contract teacher	0.217 (0.026)	0.328 (0.040)	0.257 (0.022)	-0.111**
Sekolah Penggerak	0.045 (0.013)	0.015 (0.010)	0.034 (0.009)	0.030
Survey data				
School Principal	0.111 (0.020)	0.036 (0.016)	0.084 (0.014)	0.074**
Masters degree	0.066 (0.016)	0.044 (0.018)	0.058 (0.012)	0.022
Teacher certificate	0.844 (0.023)	0.686 (0.040)	0.787 (0.021)	0.158***
Age	37.971 (0.376)	37.883 (0.490)	37.940 (0.298)	0.088
Years experience	15.922 (0.327)	15.657 (0.432)	15.827 (0.260)	0.265
PGP—batch 5	1.000 (0.000)	0.000 (0.000)	0.640 (0.025)	N/A
PGP—batch 7 or 8	0.000 (0.000)	0.015 (0.010)	0.005 (0.004)	-0.015*
Organisasi Penggerak	0.262 (0.028)	0.226 (0.036)	0.249 (0.022)	0.036
Sekolah Penggerak	0.086 (0.018)	0.073 (0.022)	0.081 (0.014)	0.013
Kurikulum Merdeka	0.041 (0.013)	0.007 (0.007)	0.029 (0.009)	0.034*
N	244	137	381	

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.

Table A.4.2: TOC I. Self-reflection and self-development

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.70 (0.03)	0.14 (0.13)	0.03 (0.12)	-0.04 (0.09)
Maps strength	0.90 (0.06)	-0.12 (0.22)	-0.16 (0.21)	0.01 (0.16)
Show strength mapping document	0.20 (0.07)	0.56** (0.28)	0.29 (0.24)	-0.06 (0.18)
Can mention are for improvement	1.00 (.)	-0.03 (0.05)	-0.05 (0.05)	-0.07 (0.05)
Observations	31	89	127	350

Notes: Row headings denote the dependent variable. x refers to running variable, centered at zero. The first column is the mean for non-PGP teachers who scored one or two points below the cutoff for selection their district. Columns (2) to (4) show the local average treatment effects, estimated using equations (1) and (2). In column (2) the sample is restricted to teachers who scored between two points below and two points above the cutoff. In column (3) the sample is expanded to teachers who scored between three below and three above the cutoff. In column (4) the same is expanded to all surveyed teachers. All estimations control for the teacher's contract type and x , interacted with treatment. Column (4) also controls for x^2 , interacted with treatment. Each regression includes additional control variables that are selected using the adaptive Lasso procedure. Baseline variables that are either predictive of treatment or predictive of the dependent variable are included. Data from the teacher survey. All dependent variables, except for the indices, are binary. "Index" is the arithmetic mean of the all the other variables in the panel. "Map strength"=1 if a teacher mentioned that they have mapped their strengths and weaknesses. "Shows strength mapping document"=1 if a teacher can show the document. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.4.3: II.1 Student-centered teaching—Positive discipline

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Panel A. Student-level				
Index	0.63 (0.01)	0.25*** (0.09)	0.15** (0.07)	0.10** (0.05)
Has class agreement (social contract)	0.85 (0.02)	0.15 (0.14)	0.07 (0.13)	0.12 (0.09)
Teacher always encourages us	0.63 (0.03)	0.20 (0.13)	0.11 (0.12)	0.06 (0.09)
Always reminded of social contract	0.43 (0.03)	0.25 (0.17)	0.18 (0.14)	0.16 (0.10)
NOT scolded by the teacher	0.62 (0.03)	0.42** (0.19)	0.19 (0.16)	0.12 (0.11)
NOT punished	0.82 (0.02)	0.22 (0.17)	0.12 (0.14)	0.11 (0.09)
No-one afraid to ask	0.43 (0.03)	0.20 (0.18)	0.22 (0.16)	0.06 (0.11)
Panel B. Teacher level				
Index	0.47 (0.06)	0.19 (0.22)	0.04 (0.19)	0.19 (0.14)
Teacher communicated reciprocally	0.70 (0.09)	0.26 (0.34)	0.26 (0.28)	0.25 (0.22)
Manager position	0.10 (0.06)	0.50** (0.23)	0.36* (0.19)	0.34** (0.16)
Motivational Provision	0.60 (0.09)	-0.18 (0.39)	-0.49 (0.33)	-0.01 (0.24)
Observations	31	89	127	350

Notes: See notes in table A.4.2. Data in Panel A is at a student level; Panel B is at a teacher level. In each case the variable “Index” is the arithmetic mean of the constituent binary indicators. See Table 5 for construction of the dependent variables.

Table A.4.4: II.1 Student-centered teaching—differentiated learning

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Panel A. Student-level				
Index	0.53 (0.02)	0.02 (0.07)	0.05 (0.08)	0.02 (0.05)
Always ask if student happy	0.56 (0.03)	0.06 (0.13)	0.09 (0.12)	0.06 (0.08)
Always help the student	0.61 (0.03)	-0.01 (0.09)	-0.05 (0.11)	0.01 (0.07)
Always provide different learning	0.54 (0.03)	-0.00 (0.11)	0.22* (0.13)	0.08 (0.09)
Assignments completed the way I like	0.36 (0.03)	-0.13 (0.12)	-0.19 (0.15)	-0.01 (0.09)
Always patient when explaining	0.46 (0.03)	-0.06 (0.11)	0.12 (0.12)	-0.13 (0.08)
Always helps when difficult with classwork	0.61 (0.03)	-0.01 (0.09)	-0.05 (0.11)	0.01 (0.07)
Always explains difficult things clearly	0.55 (0.03)	0.09 (0.12)	0.18 (0.14)	0.05 (0.09)
Panel B. Teacher-level				
Index	0.32 (0.05)	0.10 (0.18)	0.16 (0.17)	0.17 (0.13)
Adjust learning environment	0.40 (0.09)	0.15 (0.41)	-0.22 (0.33)	0.16 (0.23)
Scaffolding & independent assignments	0.47 (0.09)	0.14 (0.42)	0.26 (0.35)	0.11 (0.24)
Modify learning strategies	0.50 (0.09)	-0.26 (0.41)	-0.06 (0.35)	-0.13 (0.25)
Concrete, clear, constructive feedback	0.13 (0.06)	0.19 (0.27)	0.46* (0.24)	0.17 (0.19)
Differentiated assessment	0.10 (0.06)	0.28 (0.26)	0.38 (0.24)	0.53*** (0.19)
Observations	31	127	89	350

Notes: See notes in table A.4.2. Data in Panel A is at a student level; Panel B is at a teacher level. In each case the variable “Index” is the arithmetic mean of the constituent binary indicators. See Table 5 for construction of the dependent variables.

Table A.4.5: II.1 Student-centered teaching—socio-emotional competencies

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Panel A. Student-level				
Index	0.54 (0.01)	0.07 (0.08)	-0.01 (0.07)	-0.04 (0.05)
Teacher can be asked anything (strongly agree)	0.50 (0.03)	0.20 (0.17)	0.11 (0.14)	-0.01 (0.10)
Can tell teacher if bad day (strongly agree)	0.27 (0.03)	-0.00 (0.13)	-0.11 (0.09)	-0.09 (0.08)
Like how treated by teacher (strongly agree)	0.55 (0.03)	-0.12 (0.14)	-0.23** (0.12)	-0.13 (0.09)
Teacher always greets student with a smile	0.57 (0.03)	0.21 (0.17)	0.25* (0.14)	0.13 (0.10)
Teacher always kind to all students	0.84 (0.02)	-0.03 (0.09)	-0.13* (0.07)	-0.12** (0.06)
Panel B. Teacher-level				
Index	0.50 (0.07)	-0.74*** (0.28)	-0.35 (0.24)	-0.05 (0.17)
Empathetic	0.57 (0.09)	-0.74* (0.39)	-0.55* (0.32)	-0.45* (0.24)
Positive emotions	0.50 (0.09)	-1.20*** (0.40)	-0.71** (0.29)	-0.05 (0.22)
Socio-Emotional Content in Learning Activities	0.43 (0.09)	-0.29 (0.35)	0.22 (0.34))	0.33 (0.24)
Observations	31	89	127	350

Notes: See notes in table A.4.2. Data in Panel A is at a student level; Panel B is at a teacher level. In each case the variable “Index” is the arithmetic mean of the constituent binary indicators. See Table 5 for construction of the dependent variables.

Table A.4.6: Teach scores

	Control Mean	Impact		
	(1)	(2) $ x \leq 3$	(3) $ x \leq 4$	(4) Full
Index	-0.26 (0.19)	1.12 (0.71)	1.08* (0.60)	0.77* (0.45)
Classroom culture	-0.36 (0.20)	1.33 (0.85)	1.66** (0.67)	0.98** (0.48)
Supportive learning	3.48 (0.11)	0.66 (0.41)	0.57* (0.32)	0.75*** (0.25)
Positive behavioral	3.35 (0.12)	0.44 (0.55)	0.81* (0.44)	0.08 (0.30)
Instruction	-0.05 (0.19)	0.76 (0.74)	0.21 (0.64)	0.58 (0.46)
Lesson facilitation	3.67 (0.11)	0.67* (0.40)	0.34 (0.32)	0.16 (0.26)
Check for understanding	3.35 (0.13)	0.34 (0.54)	0.04 (0.43)	0.17 (0.32)
Feedback	2.23 (0.15)	-0.17 (0.63)	-0.16 (0.52)	0.02 (0.37)
Critical thinking	3.07 (0.13)	0.57 (0.48)	0.16 (0.45)	0.72** (0.33)
Socio-emotional	-0.20 (0.15)	0.58 (0.53)	0.73 (0.47)	0.26 (0.41)
Autonomy	2.83 (0.13)	0.23 (0.48)	0.40 (0.38)	0.33 (0.31)
Perseverance	2.25 (0.07)	0.44 (0.29)	0.29 (0.28)	0.06 (0.21)
Social collaborative	2.65 (0.15)	0.18 (0.59)	0.37 (0.48)	-0.01 (0.40)
Observations	31	89	127	350

Notes: See table A.4.2. Data from Teach classroom observations instrument. The dependent variables *Classroom culture*, *Instruction*, and *Socio-emotional* are the arithmetic means of their constituent indicators, standardized to have a mean of zero and standard deviation of one. The dependent variable *Overall* is the arithmetic mean of these three indices, also standardized to have a mean of zero and standard deviation of one.

Table A.4.7: II.1 Create communities of learning (full sample)

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.55 (0.04)	0.05 (0.13)	0.06 (0.17)	0.05 (0.09)
(a) Learning community				
Exists	0.87 (0.06)	-0.22 (0.26)	-0.04 (0.20)	-0.04 (0.16)
—initiated by teacher or school principal	0.60 (0.09)	-0.58 (0.39)	-0.62* (0.33)	-0.26 (0.25)
—initiated by teacher	0.23 (0.08)	-0.07 (0.33)	-0.27 (0.28)	-0.06 (0.19)
—initiated by teacher†	0.13 (0.06)	-0.15 (0.31)	0.07 (0.26)	-0.08 (0.19)
—peer teacher joined	0.77 (0.08)	-0.36 (0.30)	0.02 (0.24)	-0.02 (0.18)
—helps improve teaching	0.80 (0.07)	-0.07 (0.28)	0.09 (0.24)	0.04 (0.18)
(b) Self-development plan				
Maps strengths	0.80 (0.07)	0.44 (0.32)	0.28 (0.27)	0.16 (0.18)
—maps strengths (doc shown)	0.17 (0.07)	0.82*** (0.24)	0.42** (0.19)	0.02 (0.17)
—Can mention area of improvement	0.77 (0.08)	0.59* (0.34)	0.37 (0.28)	0.10 (0.20)
Asked to self-reflect as a teacher	0.60 (0.09)	0.27 (0.38)	0.20 (0.32)	0.06 (0.23)
—Initiated by PGP teacher	0.37 (0.09)	0.04 (0.38)	0.28 (0.31)	0.16 (0.23)
Asked to self-reflect on learning process	0.53 (0.09)	0.10 (0.39)	0.02 (0.33)	0.26 (0.24)
—initiated by PGP teacher	0.33 (0.09)	0.02 (0.31)	-0.16 (0.28)	0.23 (0.23)
Asked to reflect on learning process†	0.67 (0.09)	-0.21 (0.33)	-0.15 (0.30)	0.20 (0.22)
Observations	31	89	127	350

Notes: See table A.4.2. All dependent variables are constructed from the peer teacher survey, with the exception of variables denoted by †, which come from the school principal survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary.

Table A.4.8: III.1 Create communities of learning (restricted sample)

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.49 (0.06)	0.38** (0.19)	0.23 (0.15)	0.19 (0.12)
(a) Learning community				
Exists	0.81 (0.10)	-0.04 (0.28)	0.01 (0.23)	0.11 (0.19)
—initiated by teacher or school principal	0.63 (0.13)	-0.24 (0.46)	-0.36 (0.35)	-0.13 (0.28)
—initiated by teacher	0.13 (0.09)	0.33 (0.21)	0.05 (0.19)	0.29* (0.17)
—initiated by teacher†	0.06 (0.06)	-0.10 (0.26)	0.15 (0.24)	-0.17 (0.19)
—peer teacher joined	0.69 (0.12)	0.21 (0.37)	0.35 (0.28)	0.25 (0.21)
—helps improve teaching	0.69 (0.12)	0.24 (0.39)	0.34 (0.29)	0.30 (0.22)
(b) Self-development plan				
Maps strengths	0.63 (0.13)	1.07*** (0.38)	0.81*** (0.31)	0.39* (0.23)
—maps strengths (doc shown)	0.19 (0.10)	0.79*** (0.26)	0.29 (0.22)	-0.13 (0.20)
—Can mention area of improvement	0.56 (0.13)	1.40*** (0.35)	1.00*** (0.28)	0.43* (0.24)
Asked to self-reflect as a teacher	0.63 (0.13)	-0.35 (0.40)	-0.18 (0.35)	0.07 (0.26)
—Initiated by PGP teacher	0.38 (0.13)	0.32 (0.44)	0.39 (0.37)	0.25 (0.28)
Asked to self-reflect on learning process	0.56 (0.13)	0.72* (0.42)	0.13 (0.36)	0.34 (0.27)
—initiated by PGP teacher	0.31 (0.12)	0.81*** (0.25)	0.18 (0.26)	0.37 (0.24)
Asked to reflect on learning process†	0.56 (0.13)	0.18 (0.43)	0.07 (0.35)	0.29 (0.27)
Observations	16	60	83	224

Notes: See table A.4.2. All dependent variables are constructed from the peer teacher survey, with the exception of variables denoted by †, which come from the school principal survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary.

Table A.4.9: III.2 Coaching (full sample)

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.48 (0.05)	-0.02 (0.19)	-0.10 (0.22)	0.18 (0.14)
(a) Asked to discuss student-centered learning	0.67 (0.09)	0.49 (0.37)	0.43 (0.29)	0.29 (0.22)
—initiated by PGP teacher	0.50 (0.09)	0.03 (0.41)	0.08 (0.33)	0.24 (0.24)
(b) Maps out student needs	0.53 (0.09)	-0.13 (0.37)	0.04 (0.31)	0.35 (0.24)
—initiated by PGP teacher	0.30 (0.09)	-0.06 (0.38)	-0.03 (0.32)	0.35 (0.24)
(c) Observed teaching	0.80 (0.07)	-0.33 (0.29)	-0.21 (0.26)	-0.14 (0.20)
Observed by PGP teacher	0.10 (0.06)	-0.57* (0.29)	-0.40* (0.24)	-0.01 (0.16)
Observations	31	89	127	350

Notes: See table A.4.2. All dependent variables are constructed from the peer teacher survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary.

Table A.4.10: III.2 Coaching (restricted sample)

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.44 (0.06)	0.51** (0.20)	0.35* (0.18)	0.31** (0.14)
(a) Asked to discuss student-centered learning	0.63 (0.13)	0.77** (0.39)	0.52 (0.34)	0.34 (0.26)
—initiated by PGP teacher	0.50 (0.13)	0.77* (0.40)	0.32 (0.35)	0.10 (0.28)
(b) Maps out student needs	0.44 (0.13)	0.37 (0.43)	0.57 (0.37)	0.59** (0.27)
—initiated by PGP teacher	0.25 (0.11)	0.71** (0.36)	0.53* (0.31)	0.59** (0.24)
(c) Observed teaching	0.81 (0.10)	0.22 (0.30)	0.08 (0.26)	0.04 (0.23)
Observed by PGP teacher	0.00 (.)	0.20 (0.15)	0.11 (0.14)	0.22** (0.10)
Observations	16	60	83	224

Notes: See table A.4.2. All dependent variables are constructed from the peer teacher survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary.

Table A.4.11: IV.2 Vision of student-centered school development (full sample)

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.618 (0.05)	-0.12 (0.16)	0.11 (0.14)	0.19 (0.12)
Peer teacher knows vision	0.97 (0.03)	-0.03 (0.06)	0.06 (0.07)	-0.02 (0.06)
—determined by joint discussion and reflection	0.77 (0.08)	-0.31 (0.26)	0.13 (0.23)	0.31* (0.18)
—initiated by PGP teacher	0.23 (0.08)	0.07 (0.35)	0.17 (0.28)	0.17 (0.21)
—initiated by PGP teacher†	0.43 (0.09)	0.09 (0.30)	-0.06 (0.27)	0.12 (0.22)
—Teachers and school community involved	0.63 (0.09)	-0.45 (0.41)	0.02 (0.33)	0.31 (0.24)
Observations	31	89	127	350

Notes: See table A.4.2. All dependent variables are constructed from the peer teacher survey, with the exception of variables denoted by †, which come from the school principal survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary.

Table A.4.12: IV.2 Vision of student-centered school development (restricted sample)

	Control Mean	Impact		
	(1) $-2 \leq x < 0$	(2) $ x \leq 2$	(3) $ x \leq 3$	(4) Full sample
Index	0.44 (0.05)	0.27 (0.17)	0.15 (0.15)	0.25** (0.13)
—determined by joint discussion and reflection	0.75 (0.11)	0.25 (0.34)	0.22 (0.30)	0.49** (0.21)
—initiated by PGP teacher	0.25 (0.11)	0.40 (0.33)	0.11 (0.30)	0.17 (0.24)
—initiated by PGP teacher†	0.31 (0.12)	0.86** (0.34)	0.51* (0.29)	0.37 (0.24)
—Teachers and school community involved	0.69 (0.12)	-0.15 (0.41)	-0.08 (0.34)	0.22 (0.25)
Observations	16	60	83	224

Notes: See table A.4.2. All dependent variables are constructed from the peer teacher survey, with the exception of variables denoted by †, which come from the school principal survey. Index is the arithmetic mean of all the other dependent variables shown in the table. All variables are binary.

Table A.4.13: Enjoyment of teaching

	Control Mean	Impact		
	(1)	(2) $ x \leq 3$	(3) $ x \leq 4$	(4) Full
Overall	-0.38 (0.15)	2.30*** (0.81)	1.94*** (0.67)	1.22*** (0.47)
I generally enjoy teaching	4.97 (0.03)	-0.12 (0.14)	-0.13 (0.13)	0.08 (0.13)
Gladly prepare and teach my lessons	4.77 (0.08)	-0.27 (0.32)	0.21 (0.27)	0.14 (0.20)
I often have reasons to be happy while I teach	4.57 (0.09)	1.17*** (0.40)	0.86*** (0.32)	0.36 (0.24)
I generally teach with enthusiasm	4.67 (0.09)	0.24 (0.40)	0.35 (0.31)	0.15 (0.22)
I often have reasons to be angry while I teach	2.37 (0.15)	-1.13* (0.66)	-0.70 (0.67)	-0.47 (0.44)
I often feel annoyed while teaching	2.10 (0.15)	-1.77** (0.74)	-1.35** (0.56)	-1.16*** (0.38)
Sometimes I get really mad while I teach	2.90 (0.19)	-1.13 (0.85)	-1.23* (0.75)	-0.75 (0.51)
Teaching generally frustrates me	1.37 (0.09)	-0.32 (0.37)	-0.11 (0.38)	-0.22 (0.25)
I generally feel tense and nervous while teaching	1.77 (0.11)	-1.65*** (0.56)	-1.22*** (0.44)	-0.39 (0.31)
Observations	31	89	127	350

Notes: See table A.4.2. Data come from the teacher survey. All dependent variables (except for the index) are Liker scale questions ranging from (1) “Strongly Disagree” to (5) “Strongly Agree”. Overall is the arithmetic mean, standardized to have a mean of zero and standard deviation of one.

Table A.4.14: Leadership confidence and aspirations

	Control Mean	Impact		
	(1)	(2) $ x \leq 3$	(3) $ x \leq 4$	(4) Full
Panel A. Leadership ambitions				
Became a school principal	0.06 (0.04)	0.44** (0.17)	0.14 (0.11)	0.02 (0.09)
Plan to be a school principal or inspector	0.19 (0.07)	0.66** (0.33)	0.84*** (0.26)	0.50*** (0.19)
Panel B. Self Confidence as a Leader				
Index	-0.70 (0.23)	2.64** (1.07)	2.29*** (0.85)	1.61*** (0.57)
I have necessary ability to become a good leader of this school	3.68 (0.16)	1.26* (0.67)	1.38** (0.54)	0.66* (0.38)
I know what is needed to help teachers perform their duties	4.06 (0.11)	0.89 (0.56)	0.58 (0.43)	0.46 (0.30)
I believe in my ability to influence teachers	3.71 (0.14)	1.54** (0.67)	1.33** (0.54)	0.80** (0.36)
I know how to encourage teachers	3.84 (0.15)	1.52** (0.68)	1.22** (0.53)	0.70* (0.37)
Trainings have equipped me to become a school principal	3.10 (0.19)	1.63** (0.79)	1.57** (0.64)	1.55*** (0.49)
Observations	32	95	135	381

Notes: See table A.4.2. Data come from the teacher survey. The dependent variables in panel A are binary. All dependent variables (except for the index) in Panel B are Likert scale questions ranging from (1) “Strongly Disagree” to (5) “Strongly Agree”. Index is the arithmetic mean of these indicators.

Table A.4.15: Leadership confidence and aspirations (by gender)

	Impact ($ x < 3$)		Impact ($ x < 4$)	
	(1) Male	(2) Female	(3) Male	(4) female
Panel A. Leadership ambitions				
Became a school principal	3.23 (2.47)	0.36** (0.17)	-0.08 (0.36)	0.20 (0.12)
Plan to be a school principal or inspector	4.27 (3.53)	0.64** (0.27)	1.15* (0.67)	0.71*** (0.24)
Panel B. Self Confidence as a Leader				
Index — self-confidence as a leader	24.01 (30.41)	1.37** (0.68)	6.32** (2.71)	1.49** (0.65)
I have necessary ability to become a good leader of this school	6.27*** (1.75)	0.72 (0.63)	2.90* (1.56)	0.92* (0.50)
I know what is needed to help teachers perform their duties	16.23 (50.05)	0.64 (0.45)	1.98 (1.56)	0.48 (0.44)
I believe in my ability to influence teachers	35.12 (121.02)	0.75 (0.47)	5.49* (2.89)	0.73 (0.47)
I know how to encourage teachers	13.71 (18.11)	0.55 (0.40)	4.57** (1.98)	0.56 (0.38)
Trainings have equiped me to become a school principal	13.96 (19.33)	1.03 (0.80)	2.51 (1.56)	1.29* (0.72)
Observations	20	75	33	102

Notes: See table A.4.2. Data come from the teacher survey. The dependent variables in panel A are binary. All dependent variables (except for the index) in Panel B are Likert scale questions ranging from (1) “Strongly Disagree” to (5) “Strongly Agree”. Index is the arithmetic mean of these indicators.

Table A.4.16: School Principal's Perception of the Teacher's Leadership Ability

	Control Mean	Impact		
	(1)	(2) $ x \leq 3$	(3) $ x \leq 4$	(4) Full
Overall	4.77 (0.05)	-0.44 (0.27)	-0.55** (0.24)	-0.48*** (0.18)
Know how to be a good leader	4.80 (0.07)	-0.28 (0.40)	-0.43 (0.38)	-0.25 (0.25)
Helps team accomplisih tasks	4.87 (0.06)	-0.43 (0.28)	-0.69** (0.29)	-0.63*** (0.21)
Ability to influence and lead teacher	4.77 (0.08)	-0.70* (0.37)	-0.87*** (0.32)	-1.02*** (0.26)
Encourage good team	4.63 (0.10)	-0.34 (0.45)	-0.19 (0.35)	-0.11 (0.25)
Observations	31	89	127	350

Notes: See table A.4.2. Data come from the school principal survey. All dependent variables (except for the index) are Likert scale questions ranging from (1) “Strongly Disagree” to (5) “Strongly Agree”. The school principal was thinking about the leadership abilities of the target teacher when answering the questions. Overall is the arithmetic mean of these indicators.



THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP



CWPF
CHINA-WORLD BANK GROUP
PARTNERSHIP FACILITY



Australian Government