RISE Working Paper 22/106
August 2022

Is the Game Worth the Candle? Examining the Effectiveness of Initial Teacher Education in Indonesia

Asri Yusrina, Ulfah Alifia, Shintia Revina, Rezanti Putri Pramana, and Luhur Bima

Abstract

An impactful teacher education programme equips teachers with knowledge and skills to improve their effectiveness. Empirical findings on the effectiveness of teacher preparation programmes show that the accountability of institutions and teachers should not only be based on the knowledge or skills produced but also on student learning. Our study aims to evaluate the effectiveness of a pre-service teacher education programme in Indonesia, known as *Pendidikan Profesi Guru Prajabatan* or PPG. PPG is a one-year full-time programme in addition to four years of undergraduate teacher education (Bachelor of Education). PPG graduate teachers pass a selection process and receive a teaching certificate upon completion of the programme. We use mixed methods to understand the differences in the outcome of PPG graduates majoring in primary school teacher education to their counterparts who did not attend PPG. To estimate the impact of PPG, we exploit the combination of rules and events in the selection process which allows us to estimate the impact of PPG on teacher performance using fuzzy regression discontinuity design (RDD). Once we attest to the validity of the fuzzy RDD, we find that PPG has no impact on a teacher's professional knowledge and student outcomes in numeracy and literacy. We argue that this is due to the ineffective selection mechanism in distinguishing the PPG and the comparison group. We conclude that as an initial teacher training programme, PPG did not improve teacher effectiveness. Despite incorporating best practices from effective teacher training into the programme design, PPG does not appear capable of producing a higher-quality teacher.

Keywords: teacher professional education, programme evaluation, content knowledge, pedagogical content knowledge, student learning outcomes

Is the Game Worth the Candle? Examining the Effectiveness of Initial Teacher Education in Indonesia

Asri Yusrina
SMERU Research Institute

Ulfah Alifia
SMERU Research Institute

Shintia Revina
SMERU Research Institute

Rezanti Putri Pramana
SMERU Research Institute

Luhur Bima
SMERU Research Institute

Acknowledgements:

The authors would like to thank Daniel Suryadarma, Sudarno Sumarto, and Emilie Berkhout for their comments and review of the paper.

This is one of a series of working papers from "RISE"—the large-scale education systems research programme supported by funding from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), the Australian Government's Department of Foreign Affairs and Trade (DFAT), and the Bill and Melinda Gates Foundation. The Programme is managed and implemented through a partnership between Oxford Policy Management and the Blavatnik School of Government at the University of Oxford.

Please cite this paper as:
Yusrina, A., Alifia, U., Revina, S., Putri Pramana, R. and Bima, L. 2022. Is the Game Worth the Candle? Examining the Effectiveness of Initial Teacher Education in Indonesia. RISE Working Paper Series. 22/106. https://doi.org/10.35489/BSG-RISE-WP_2022/106

Contents

List of Tables

List of Figures

1. Introduction

Access to quality teaching is considered one of the primary factors contributing to learning opportunity that leads to high learning outcomes (Canales and Maldonado, 2018; Chetty et al., 2014; Rivkin et al., 2005; Rockoff, 2004). In the school setting, it is widely acknowledged that learning occurs as a result of an interaction between individual learners and their surroundings—that is generally prepared and carried out by the teacher (Blömeke et al., 2016). Teacher quality should thus matter, and this is especially relevant in developing countries and in the primary grades where students tend to spend most of the day with a single teacher (Mincu, 2015).

Many studies indicate that the effectiveness of teachers, although defined in different ways, is an important variable influencing students' achievements. Previous studies show that teachers' years of teaching experience—although diminishing after five years (Clotfelter et al., 2007; Rivkin et al., 2005)—and teachers' test scores (D'Agostino and Powers, 2009; Clotfelter et al., 2007) consistently show higher student learning gains. However, in terms of teacher training, the results are inconclusive. Jacob and Lefgren (2004) found no significant effect of teacher training on student outcomes, while Harris and Sass (2011) found positive effects for teacher content-focused training on student outcomes but not so much on the pedagogical training.

In the developing country context, the effect of teacher training has been unclear. Teacher qualification and initial training do not significantly affect teacher performance. In Nigeria, ESSPIN research (2008) found low levels of teacher content knowledge in literacy and numeracy for the Grade 6 primary school curriculum. Overall, only 0.4% of Nigerian teachers achieved an average percentage score above 80% as required. Specifically, the level of teacher qualification did not affect test scores. Teachers with formal qualifications did not perform better than those without formal qualifications. Likewise, research in Sub-Saharan African countries also found that students whose teachers had no initial training did not perform worse than those taught by trained teachers (Bernard et al., 2004 cited in Best et al., 2018). This raises the question: If teacher preparation programmes (i.e., teacher education and training) appear to have no impact, can teaching skills be taught through education and training?

Nevertheless, many others believe it is possible to improve classroom teaching and learning practices in ways that lead to higher student learning through high-quality teacher preparation (Cochran-Smith and Maria Villegas, 2015; Tatto, 2015; Döhrmann et al., 2012). A quality teacher preparation programme provides teachers with access to knowledge that will improve their effectiveness, including general pedagogy knowledge, subject matter knowledge, knowledge about child development and learning, as well as skills to translate teaching ideas into useful learning experiences (Darling-Hammond, 2016). Although, it is widely accepted that several issues affect the degree to the effectiveness of teacher preparation (Popova et al., 2019; Darling-Hammond, 2017; Ingvarson and Rowley, 2017; Tatto, 2015; Westbrook et al., 2013). First, the focus on teaching content in the curricula is often adopted at the expense of teaching methodology. This approach is believed to be driven by the intake into teacher training programmes, with the low level of entry qualifications for teacher candidates. Second, teacher training curricula have often failed to keep pace with the real needs of classrooms. The curricula fail to resource the classroom practice element of teacher training programmes effectively. Third, graduation criteria are based largely on examinations or written assessments, not on assessing practical teaching ability. Fourth, there is a large discrepancy in the quality of teacher training providers. Teacher educators themselves may not have received professional development and are out of touch with schools.

Following the suggestions to provide high-quality teacher education, the Indonesian Government instituted a one-year post-undergraduate teacher training expected to produce highly skilled teachers, *Pendidikan Profesi Guru* (PPG) *Prajabatan* or Pre-Service Teacher Professional Education. The Government uses the completion of this programme in a high-stakes way to appoint civil servant teachers and provide teaching allowance.

The present study aims to evaluate the effectiveness of the PPG. The number of seats in PPG is limited, and not all Bachelor of Education holders can secure a place in the programme. By design, PPG aims to address the shortcoming of teacher education mentioned above. PPG has selected the best candidates through a screening exam. It focuses on the teaching methodology and equips teachers to teach effectively. PPG stresses the importance of school-based experience. The PPG assessments include a component of teaching practice. Despite the promise brought by PPG, the added value of the Programme to its graduates, who must undertake an additional year of training, remains unclear. Central to our inquiry is whether sending teachers to a one-year additional training programme with the aforementioned features, on top of a Bachelor of Education, results in a better-quality teacher.

In this study, we use a mixed-method approach to compare the competency of PPG graduates majoring in primary school teacher education (PSTE) to those of their counterparts with similar characteristics but only hold a Bachelor of Education degree. We evaluate PPG's impact using the Regression Discontinuity Design (RDD) for teachers' content knowledge (CK) and pedagogical content knowledge (PCK) in literacy and numeracy. We also assess the PPG's impact on student learning outcomes. Both PPG graduate teachers and those who did not attend PPG (non-PPG teachers) are eligible to teach in schools. The eligibility of the PPG participants or teacher candidates to join the programme was determined using a cut-off rule in the online admission test. Applicants who scored above the predetermined cutoff were eligible to participate in PPG, while those who scored below were not admitted. However, not all applicants who passed the selection participated in the Programme. In fact, some of the unsuccessful applicants took another try and were admitted the following year. The combination of rules and events in the selection process allows us to estimate the impact of PPG on teacher performance using fuzzy RDD. We also conducted in-depth interviews with selected teachers from the two groups, their principals, their students, and the students' parents, to understand qualitatively how different stakeholders define the indicators, and assess, teacher quality. Consequently, our quantitative work enables us to understand the impact of the PPG graduates' competency and student learning outcomes. Whereas the qualitative work informs us of how the existing PPG is conducted and highlights what would work to make a better initial teacher training than PPG.

We find that PPG has no impact on teachers' professional knowledge as measured by standardised scores in CK and PCK tests. We also find no impact on test-score performance in numeracy and literacy of students taught by teachers who graduated from PPG. Given the conditions of this quasi-experiment study, we find that the treatment group is different from the comparison group in several characteristics. Since we account for this confounding characteristic in the data, our method robustly yields the local average treatment effect on the compliers, i.e., the PPG graduates, at the cutoff, as explained in Frölich and Huber (2018). Potentially, the estimates have low power. We argue that it is attributed to inclusion and exclusion error of defining quality teacher resulting from an ineffective selection mechanism.

Furthermore, as the Indonesian Government established PPG in such a way that fuzzy RDD allows identification of the impact of PPG on teacher effectiveness, the findings from the present study have two significances. First, our findings showed that if there is a way to conduct initial teacher training that

improves teacher practices that lead to higher student outcomes, it is not PPG. Second, having included best practices from effective teacher training in the design, PPG appears capable of producing higher-quality teachers when it is likely not. In short, PPG process, at both design and implementation level, has explained the lack of its graduate's quality. While our findings may support the assumption that teaching skills might not be developed through training such as PPG, our reflections that compare PPG with teacher training programs with effective features suggested by literature indicate that there exists initial teacher training program - designed differently from how PPG was set up – that can serve that purpose.

We structure the paper into five sections. The next section provides an overview of the PPG programme features. The third section describes the research methods. The fourth section presents the results and findings of both quantitative and qualitative studies. The fifth section discusses a broader analysis of how PPG features link to the results presented in the fourth section.

2. The Pre-Service PPG Programme and Initial Teacher Education in Indonesia

In Indonesia, the entrance into the teaching profession is through undergraduate teacher education. There are over 400 teacher colleges or universities known as LPTK (*Lembaga Pendidikan Tenaga Kependidikan*) across Indonesia that offer a four-year undergraduate programme in education. LPTK graduates hold a Bachelor of Education (B.Ed) degree. According to Worldbank (2010), only half of these LPTK graduates will enter the teaching profession. The B.Ed is considered a concurrent model of teacher education because it includes pedagogical and content knowledge courses that take place within the four years of education. This degree has been subject to criticism partly because of its low entry requirements and its less school-based teaching experience (Worldbank, 2010).

To answer the said criticism of initial teacher education, Indonesia reformed its teacher training. The pre-service PPG is a one-year full-time teacher training programme that employs a centralised screening process and emphasises the importance of school-based experience. The pre-service PPG is considered a consecutive teacher training model because it targets teacher candidates who have completed an undergraduate degree in a related discipline. This pre-service training programme only applies to undergraduates with less than five years of teaching experience. The years of teaching experience differentiate teacher candidates from experienced teachers in the existing force who are directed to pursue in-service PPG.

In addition to more selective teacher candidates, the pre-service PPG can only be managed by LPTKs selected by the Ministry of Education and Culture (MoEC). In 2018, 43 LPTKs (of over 400 LPTKs registered in Indonesia) were eligible to host PPG, and from that list, only 29 LPTKs were allowed to manage pre-service PPG for PSTE. With the selected number of LPTKs, the available pre-service PPG seats are limited. Such a quota was set because, in 2018, the programme was provided in the form of scholarship, therefore subject to fiscal availability.

The pre-service PPG specifically attracts aspiring novice teachers with undergraduate degrees from the related discipline to apply for this full-time teacher training as it offers certification upon completion of the programme. Holding a teacher certificate leads to a higher probability of gaining a higher income—earning double their base salary (De Ree et al., 2018)—although, for novice teachers, such likelihood also depends on their employment status. Prior to 2018, teacher certification was only applicable to

experienced teachers in the existing force. The certification process for the said teachers has changed over time, from only portfolio review to including a short period of in-service training using blended mode. The in-service training is known as PPG *dalam Jabatan* or PPG for in-service teachers. Currently, the pre-service PPG for novice teachers and in-service PPG for the existing teaching force are also pathways to obtaining teacher certificates. In this study, we focus on the pre-service PPG—hereafter, we use "PPG" interchangeably with "pre-service PPG" and exclude analysis of the in-service PPG.

The pre-service PPG graduates are a pool of new teachers coming into the profession that meets the requirement of the 2005 Teacher Law. The Law states that teachers in Indonesia must hold a bachelor's degree in the subject they teach and a teaching certificate. According to the Law, only such teachers are supposedly allowed to teach. However, schools need teachers in their classrooms, and in the actual job market, teachers are only required to hold a bachelor's degree, and then they can start teaching.

Apart from the certification, since pre-service PPG employs a screening process, the admitted teacher candidates are considered the best relative to other pools of new teachers. They have passed the admission requirements, such as a minimum grade point average of 3 (out of 4.00 scale) from an accredited university. Once they meet the admission criteria, they must pass the centralised, standardised online tests administered by the MoEC. The tests measure academic aptitude, English skills, and pedagogical knowledge. Those who pass the test must pass an interview by LPTKs using instruments developed by the MoEC. Acceptance into the programme is based on the final score, the weighted average of the online admission test scores, and the interview score.

The one-year full-time training programme was structured into one semester of in-class sessions and one semester of school placement and action research. According to the PPG curriculum and guidelines developed by the MoEC, during the first semester, teacher candidates in the PSTE major learn in small groups how to prepare lesson plans, develop students' worksheets, use appropriate teaching aids, and develop the corresponding assessments for five subjects taught in Grades 1 to 6 primary school (math, language, natural science, social science, and civic education). The candidates must also perform peer teaching on selected lessons during this period. In the second semester, the teacher candidates are placed at partner schools in groups of five, supervised by a teacher educator and a mentor teacher. For each teacher candidate, PPG requires a minimum of eight classroom observation visits by both teacher educator and mentor teacher. During this school-based activity, candidates must also write an action research report.

Upon completing the coursework and school placement, PPG participants must take exit exams consisting of two components. The first component is a computer-based exam on teachers' general pedagogical knowledge as well as their CK and PCK on the five main subjects taught in primary school. The second component is a microteaching exam, where a teacher candidate is observed by three external examiners (the teacher educator supervisor and mentor teacher are not part of the panel). Teacher candidates can choose to teach two hours of lessons from one, or more, of the five main subjects taught in primary school. The results from the two exams will determine whether the teacher passes the Programme. Failing teachers have up to two times to retake the tests. From our data, all PPG participants eventually passed the examination.

3. Methods

We conducted a mixed-methods study, which involved collecting and analysing longitudinal quantitative and qualitative data from October 2018 to June 2021 to answer the research questions. This unified approach aims to connect what teacher candidates experienced when they underwent PPG with their performance after completing the Programme and the ultimate outcomes. The qualitative study also serves to explain some results from the quantitative survey.

3.1. Quantitative Method

3.1.1. Data and Sample

To estimate the effect of PPG, we took advantage of the PPG selection process. Teacher candidates' eligibility to participate in PPG was made using a cutoff rule in the online admission test. Applicants who scored above the predetermined cutoff were eligible to participate in PPG, while those who scored below were denied. We used a dataset from the PPG pre-service selection for primary school teachers in 2018 to specify the study sample. We have administrative data of 4,339 applicants; 1,291 passed the screening and were enrolled in the 2018 PPG hosted by twenty-nine LPTKs.

Since no secondary data on outcome variables (which could measure teacher performance) were available, we collected primary data. To manage logistical issues in the data collection, we selected seven LPTKs in Java. This selection is because twenty-two out of the twenty-nine LPTKs that held pre-service PPG were in Java. We purposively chose the seven LPTKs because these colleges had larger numbers of teacher candidates relative to other LPTKs in Java that held pre-service PPG. We also assumed that PPG graduates would find teaching positions in the surrounding districts of their teacher colleges.

To control the quality of the teacher colleges and applicants' education backgrounds in the targeted sample, we selected applicants with undergraduate degrees from the seven teacher colleges. As for the treatment group, the applicants' undergraduate universities were the same as the teacher colleges where they attended PPG. We set it up in such a way to ensure the background characteristics of the sample are similar.

Based on the administrative data of the seven LPTKs, we identified 688 applicants (491 were admitted into PPG and 197 were not because they failed the screening test). From October to November 2018, we visited six LPTKs and surveyed 316 PPG participants. The following year, we carried out a phone survey to track the applicants' whereabouts, whether they worked as teachers, and the location of the schools where they taught. We contacted 291 teachers via phone but could only visit 197 teachers (see Appendix Table A1). The teachers were scattered in sixty-seven districts in Java. We were able to locate 122 teachers who participated in PPG—their scores were above the cutoff in the online admission test. Meanwhile, the comparison group comprised seventy-five teachers who could not enrol in PPG (the non-PPG teachers) because their admission scores were below the eligibility cutoff.

We collected three outcome variables from our sample teachers. Empirical findings on the effectiveness of teacher professional development programmes show that the accountability of institutions and teachers should not only be based on the knowledge or skills produced (Kleickmann et al., 2013) but also on student learning (Goldhaber et al., 2013; Harris and Sass, 2011; Boyd et al., 2009). Accordingly,

we administered instruments designed to assess and predict teacher effectiveness (i.e., teachers' CK and PCK scores) and students' test scores.

The 2018 Pre-Service PPG guideline aimed to produce professional teachers. Since professional knowledge is key to teachers' professional performance (Shulman, 1986), we assessed teachers' CK and PCK as performance indicators. Kleickmann et al. (2013) explained that teachers might acquire their CK and PCK through three different periods in their careers: pre-service teacher education, induction phase, and in-service phase. Each phase does not necessarily offer certain aspects of CK and PCK, but creates learning opportunities to develop CK and PCK. Similarly, teachers who graduated from PPG had longer learning opportunities to develop their CK and PCK through workshops, lesson observations, and teaching practices in partner schools during the programme. Thus, we hypothesised that PPG had an impact on teachers' CK and PCK.

Teachers are central in promoting student achievement. Research on the effect of teacher preparation programmes on student learning outcomes aims to evaluate the programme's effectiveness (Goldhaber et al., 2013; Harris and Sass, 2011; Boyd et al., 2009). By doing so, the teacher preparation programme is to be held accountable for the extent to which its graduates perform in the classroom.

Boyd et al. (2009) found that programmes that focus more on classroom practice in an actual setting are positively associated with student learning in reading and math during a teacher's first year of teaching. Meanwhile, programmes that focus on content preparation showed an effect on student learning in the teacher's second year of teaching. Contrastingly, Harris and Sass (2011) found that teacher preparation programme was unrelated to students' outcomes. Goldhaber et al. (2013) noted that the indicators of training programmes in twenty-one teacher colleges in the United States, on average, only predicted student achievement in reading but not math. When we visited our sample teachers in February 2021, two years after the 2018 pre-service PPG ended, we sought to examine whether the programme had any effect on student learning outcomes.

3.1.2. Descriptive Statistics of Sample

The descriptive statistics on the observable covariates for both treatment and comparison groups are presented in Table 1. We see that the compositions of gender, age, and teaching experience did not differ significantly between the two groups. However, the average GPA of teachers who graduated from PPG—from the entire sample—was higher than that of non-PPG teachers.

To capture school-level data, we generated a school quality index from the 2017 Ministry of Education's Education Data Centre (*Data Pokok Pendidikan-Kebudayaan*/DAPODIK[1] 2017). One may assume that PPG teachers are sorted to good schools, defined here as the higher quality index. However, the recruitment process for civil servant teachers is carried out in such a way that teachers can only apply to schools with vacant teaching positions. Teachers are likely to be accepted if they do not select a well-performing school but rather a school with a vacant teaching position. This is in line with our sample

[1] The DAPODIK includes information on the school's infrastructure, staffing, accreditation, and the number of students at the school level. The index was a composite of the following indicators: teacher-student ratio, number of teachers with undergraduate degrees, characteristics of classrooms with good quality, the provision of library, the provision of science and computer laboratories, and school's accreditation status. DAPODIK is only available for schools managed by the MoEC. Thus, for teachers teaching in Islamic schools called madrasa, which are under the Ministry of Religion Affairs, we treat them as the missing data on the school quality index.

teachers' answers when asked in the teacher questionnaire about their reasons for choosing their current school (see Appendix Table A2). On that note, we can see in Table 1 that the school's quality related to their PPG graduate teachers and non-PPG teachers, from the complete sample, are not substantially different.

Table 1 Descriptive Statistics on the Observable Covariates (n=197)

Variables	Teachers Graduated from PPG					Non-PPG Teachers				
	n	Mean	sd	[95% Conf. Interval]		n	Mean	sd	[95% Conf. Interval]	
Gender	122	0.79	0.41	0.71	0.86	75	0.75	0.44	0.65	0.85
Age	122	24.48	1.23	24.26	24.70	75	24.72	1.38	24.40	25.04
Undergraduate GPA	122	3.61	0.17	3.58	3.64	75	3.51	0.15	3.47	3.54
Years of teaching experience prior to PPG	119	2.11	1.16	1.90	2.32	73	2.56	1.14	2.30	2.83
School quality index	120	0.09	1.64	-0.21	0.38	72	-0.16	1.33	-0.47	0.16

3.1.3. Regression Discontinuity Design Method

The PPG eligibility depends on the online admission test. Applicants whose online admission test scores were above the predetermined cutoff could participate in the Programme, while those who scored below could not join PPG. Under the PPG entry selection rule, applicants who scored above 50.00 in the online test proceeded to the interview stage, while those whose scores were below 50.00 were automatically denied. Applicants who continued to the interview stage received their final scores, and those with scores above 60 were eligible to participate in PPG.

Both cutoffs are independent of the teacher candidates as they are determined by the MoEC. Since there are two cutoffs, we employed a pooling strategy and only used the online admission test scores as a running variable. If we used the final scores as our running variable, only applicants who passed the online test but failed the interview would be eligible for the comparison group.

In addition to that, the eligibility rules were not enforced strictly. Some eligible applicants opted out of PPG, namely the non-compliers. We also found that some unsuccessful applicants in our sample had undertaken the Programme the following year by the time of our visit. Figure 1 is a graph plotting the probability of receiving treatment as a function of the running variable. It shows that there is incomplete compliance on both sides of the groups.

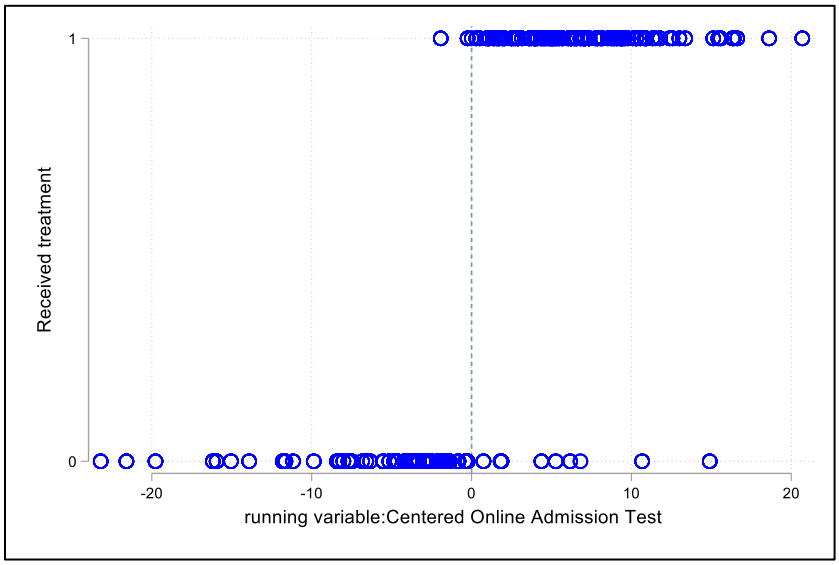

Figure 1. Treatment Status based on the Online Admission Test

The combination of programme eligibility assigned discontinuously based on a score and a cutoff and weak compliance with eligibility status allowed us to estimate the impact of PPG on teacher performance using fuzzy RDD.

Let T be an indicator of assignment to Treatment or the result of the selection when applicants are admitted into PPG

$$T_i = \begin{cases} 1 & X_i \geq c \\ 0 & X_i < c \end{cases}$$

Let D be to denote whether the treatment was received by unit *i*, that is when the applicants enrolled and participated in PPG. In the fuzzy RDD case: $D_i \neq T_i$.

The first-stage estimation:

$$D_i = \delta_0 + \delta_1 X_i + \delta_2 T_i + \eta_i \qquad (1)$$

In fuzzy RDD, D_i is endogenous and $T_i = 1_{(X_i \geq c)}$ is the instrument. The probabilities of receiving treatment of individuals just above and below the cutoff are different because of crossover or no show.

The second-stage estimation:

$$Y_i = \alpha + \beta X_i + \tau \widehat{D}_i + \epsilon_i \qquad (2)$$

It means T_i affects Y_i only through D_i at the cutoff.

The fuzzy RDD reduced form is:

$$Y_i = \alpha_R + \beta_R X_i + \tau_R T_i + \epsilon_i \qquad (3)$$

As in the first stage, we clustered standard errors at levels of the running variable, i.e., teacher candidates' level. Our coefficient of interest is τ_R, which represents the causal effect of participating in PPG for a teacher candidate who fell just above of the required passing score on online admission test score compared to candidates who scored just below the threshold. Three different outcome variables of individual teacher candidate (Y_i) were investigated: their CK score, their PCK score, and their students' test scores. All outcome variables focus on two subjects: numeracy and literacy.

We also constructed teacher background variables from the teacher questionnaire, covering age, gender, undergraduate GPA, and years of teaching experience before joining PPG. Student background variables were constructed based on data from a parent questionnaire, including gender, age, parents' education, housing quality index, and whether or not the students took private lessons. The housing quality index was defined as a composite of seven indicators, including the material used for roof, wall, and floor; access to drinking water; sanitation; electricity; and fuel for cooking. For the school background variables, we used the school quality index. To address the COVID-19 pandemic situation, we also included variables of the number of days per week the teachers teach. During the data collection, the health protocol for schools to implement in-person learning was 50% but it also depended on the schools in terms of the number of days per week.

Following Imbens and Kalyanaraman (2012), we modelled the relationship between online admission test scores and the candidates' outcomes as "locally linear". It means that the RDD estimated the average effect of the sub-population with a running variable, which in our case is the online admission test score, close to the cutoff.

We chose an optimal bandwidth using the mean squared error method and fit models across additional bandwidth choices. Using wider bandwidths allowed us to increase the statistical power and precision. In the threats to validity, we performed several tests to verify that the results were not sensitive to the bandwidth choices or the functional form of the forcing (running) variable.

All regression discontinuities have the potential to be undermined by failures of important assumptions. Following Cattaneo et al. (2018), we first investigated whether any evidence existed to suggest manipulation testing of the running variable to test for self-selection or sorting. The idea was that supposed candidates knew the cutoff point and wanted to manipulate the test score or the assignment variable to receive the treatment. Since the 2018 PPG was subsidised by the government (similar to scholarship), teacher candidates had the motive to get accepted. Manipulation of the teacher candidate's position relative to the cutoff was highly implausible. For instance, candidates could not manipulate their positions to be accepted into PPG because the online admission test score was administered centrally at the national level. If candidates cannot manipulate the running variable, the variation in treatment is similar to being randomised. Whether the candidates passed the online admission cutoff or not, it was by chance. Thus, candidates who were just above and below the cutoffs are likely similar. Despite the absence of a real threat to the validity of the assignment variable, the manipulation testing plot (see Appendix Figure A1), shows a jump around the discontinuity used to assign candidates to PPG. However, the RD manipulation testing rejected the null hypothesis of manipulation around the cutoff.

The assignment variable was used to assign units to treatment, so differences in outcomes should not be attributable to other potential mechanisms, particularly from pre-intervention covariates. For another

check on the appropriateness of the RD approach, we examined the distribution of the covariates we used as control variables to ensure that there were no other discontinuities that might have generated the results. To examine potential discontinuities, we checked the relationships between the running variable and the observable covariates at the teacher- and student-level data.

For the teacher-level data, we checked whether there were discontinuities in the observable characteristics at the cutoff. These covariates include age, standardised undergraduate GPAs, and years of teaching experience prior to admission to PPG (see Appendix Figure A2). We found that there was a potential for undergraduate GPA to influence differences in the outcome. This is likely to be the selection effect mentioned in Harris and Sass (2011) and Goldhaber et al. (2013). Such that the PPG effect may be influenced by teacher candidates' pre-intervention ability correlated with selection into the programme. Thus, we need to isolate the programme's effect from the selection effect by including undergraduate GPAs as a control for pre-intervention ability. As the sample teachers are further away from the cutoff, there are significant differences among PPG graduates (see Appendix Table A3).

We also checked discontinuities at the cutoff of observable covariates pre-intervention at student-level data against the running variable. We checked students' housing quality index (i.e., as a proxy of socioeconomic status), parents' education attainment and teaching frequency during Covid-19. We found no discontinuities on these covariates except for the meeting frequency (see Appendix Figure A3).

An important caveat of our study is that—as in any RDD—our analysis applies only to teachers at the cutoff. Furthermore, these teachers were only representatives of the teachers enrolled in the pre-service PPG and were graduates of the seven teacher colleges.

3.2. Qualitative Method

3.2.1. Participant Selection and Data Collection Procedures

The qualitative approach aims to complement the quantitative findings by elaborating on the features of the existing PPG to the effective components of teacher education programmes suggested in the literature. This elucidates why PPG has not turned into a teacher education programme that can effectively develop teachers' skills and competencies, as initially found by our quantitative results.

We conducted in-depth interviews with selected PPG participants at two different points of time. The first was when the teacher candidates participated in PPG courses in 2018. We also interviewed representatives of PPG instructors from LPTKs and partner schools to understand how the curriculum and instructors were prepared and how the Programme was implemented in different institutions. The second was when the teachers had graduated from PPG and had been teaching for about a year or two in 2020 and 2021. The data collection in 2020 was only partially completed due to the COVID-19 pandemic. We continued the data collection in 2021. With the teachers' consent, we had the opportunities to talk with some parents and the school principals.

We purposively selected twenty-nine teacher respondents, with twenty teachers enrolled in the 2018 PPG, while the nine failed to enter the Programme (the latter being the comparison group). The respondents were selected after the first quantitative survey was completed in 2018. To capture variations in the respondents' perspectives, we selected the study participants based on four criteria: gender, undergraduate GPA, classroom experience, and motivation to join PPG. Below are the rationales of our selection criteria:

1. Based on DAPODIK 2017, 63% of primary school teachers are female. In the qualitative study, we also purposively selected more female teachers (sixteen) than male teachers (thirteen) to reflect the overall proportion of teachers based on the gender data.

2. GPA may influence teacher performance. Two-thirds of the selected PPG participants had GPAs above 3.50. In the qualitative study, we decided on respondents who represent both moderate GPA (3.00–3.40) and high GPA (3.50–4.00) to understand how teachers with different GPAs perceive teaching qualities and whether stakeholders see that a teacher's GPA influences their performance.

3. Novice teachers enrolled in pre-service PPG have zero to five years of teaching experience. According to Giallo and Little (2003), having classroom experience will increase the confidence of new teachers in managing a classroom. We assumed that having more classroom experiences would help novice teachers perform better during their study and adapt more easily to their new environment (school). To better understand this assumption, we selected respondents with and without classroom experiences.

4. The quantitative survey asked one question on teacher motivation to join the PPG. We sought to understand the variations in the twenty PPG graduate teachers related to their initial motivation of joining PPG and whether their motivation influences their behaviour and practice in teaching.

3.2.2. Data Analysis

All data collected were documented and developed into field notes. We then performed open coding—breaking down the data into specific categories—through NVivo. Generally, we closely examined the data and synthesised it to find themes, patterns, and relationships.

As we learn from global literature, it is possible to establish a teacher training programme that can improve classroom teaching and learning practices. We elaborate features of effective teacher training programme that has been included in PPG. By doing so, we aim to explain the extents to which the programme affects the quality of its graduates.

4. Results

The results of this study are presented in three parts to ensure clarity between the impact of PPG on teacher competency, student learning outcomes, and knowledge of the programme features responsible for building teacher effectiveness.

4.1. Impact on Teachers' CK and PCK

In this section, we hypothesised that PPG effectively increases teachers' CK in numeracy and literacy. In the fuzzy RDD, we employed a two-stage least squares (2SLS) approach to estimate the local average treatment effect (LATE) of participating in PPG. Another way to explain this local estimator is to note that as one approaches the cutoff point, the resulting treatment and control groups become increasingly similar in all ways except for receipt of treatment.

Figure 2 and Figure 2 depict the reduced form estimates for the outcomes of interest, standardised scores of teachers' CK and PCK in numeracy and literacy. Each plot represents coefficients from a separate regression for the estimation without and with covariates in numeracy and literacy. In the estimation of PPG effect on the teacher effectiveness, the number of sample teachers who completed the CK and PCK test in both treatment and control groups is 185 and 193, respectively. The number of sample teachers in each test affects the possible bandwidth used to estimate the LATE.

For the estimation without covariates, we used smaller bandwidth. The suggested smaller bandwidth results from the fact that as one approaches the cutoff point, the observable characteristics of treatment and control groups become increasingly similar in all ways except for receipt of treatment. In such a case, covariates are not needed. Under stronger identifying conditions, Calonico et al. (2016) require the covariates to be continuous at the cutoff to ensure a valid comparison group.

For estimation with a broader range of bandwidths, we included covariates of teacher's age, standardised value of undergraduate GPA, and years of teaching prior to PPG. Accounting for covariates is plausible for several reasons, such as to reduce variance of the estimated treatment effect and mitigate small sample bias (Frölich and Huber, 2018). With a smaller sample close to the cutoff point, it is also possible to include samples further from the cutoff. In the presence of covariates, Frölich and Huber (2018) explained that we can still identify local treatment effect when the covariates are likely to influence the running and the outcome variables.

In our case, from the RD plot analysis, we found that undergraduate GPA creates a discontinuity in the running variable, such that PPG graduate teachers are likely to be those with higher GPAs. Previous studies focused on teacher tests show that the GPA of the prior education predicts a teacher's CK and PCK (Kleickmann et al., 2013). Therefore, we included the standardised undergraduate GPA and assume that identification of LATE is valid conditional on GPA.

Figure 2 and Figure 3 show that the PPG's estimated effects for RD with smaller bandwidth and without any covariates are insignificant in teachers' CK in numeracy and literacy and teacher's PCK in numeracy. We found a significant positive impact of PPG on teachers' PCK in literacy. The significant effect for specification without covariates was likely confounded by differences in the undergraduate GPA and years of teaching prior to PPG between PPG graduate teachers and the non-PPG teachers. We further investigated the impact of pre-service PPG on teachers' CK and PCK using wider bandwidth and with covariates to see if those results hold.

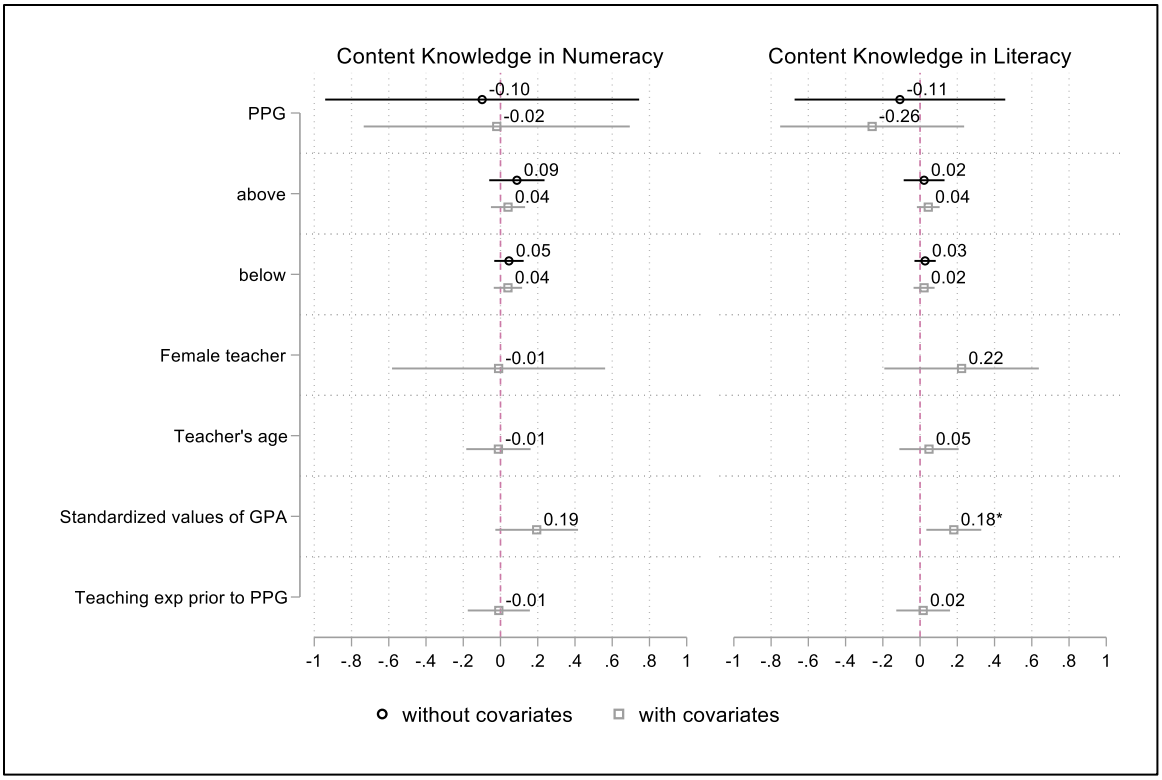

Figure 2. Estimation Results for Teachers' Content Knowledge Scores

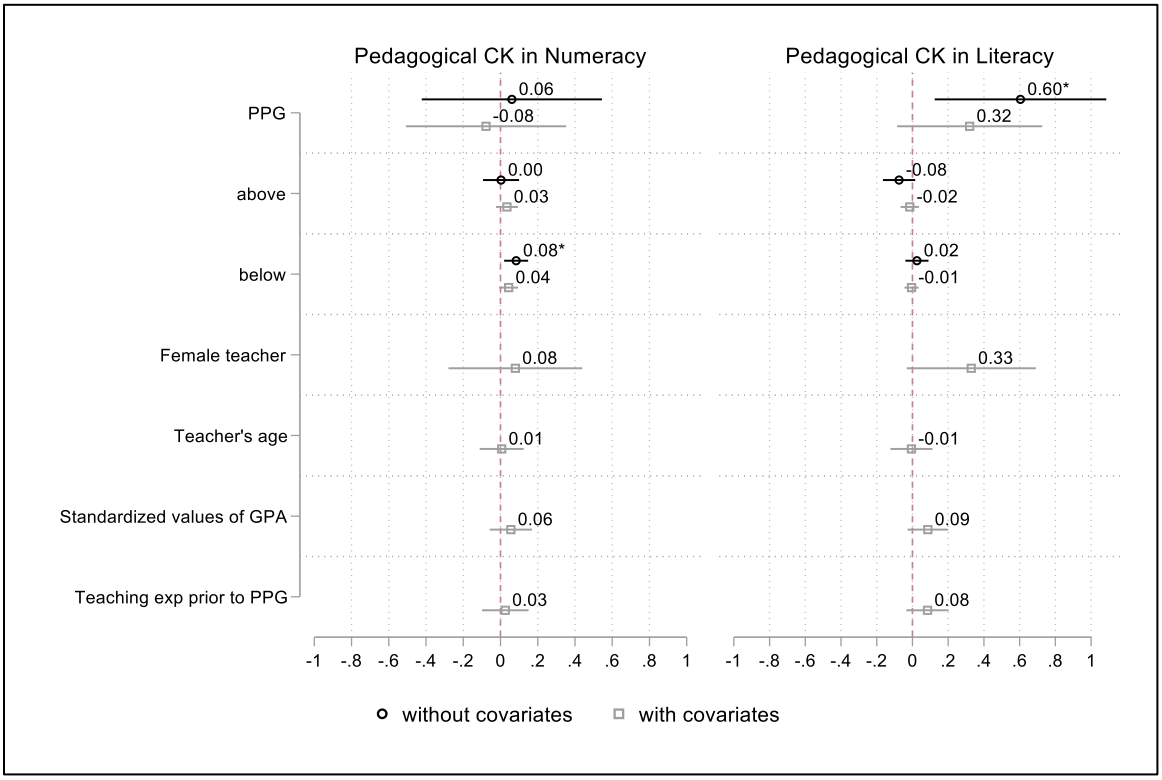

Figure 3. Estimation Results for Teachers' Pedagogical Content Knowledge Scores

When we used larger bandwidth and included covariates that we believed might be correlated with the outcome and would increase the statistical precision of impact estimates, we found that PPG had no statistically significant impact on teachers' CK and PCK scores in numeracy and literacy. We found that the significant effect of PPG on teachers' PCK in literacy diminished as more observations were included (i.e., further away from the threshold). The standard error of being a PPG graduate teacher became slightly smaller as we included more observations around the cutoff and covariates.

As part of the sensitivity analysis, we used bandwidths slightly larger than the optimal bandwidth; none of our impact estimates were statistically significant. Our preferred robust specification for the impact of PPG on teacher competency is the larger bandwidth with covariates (as presented in Figure 2 and 3). The underlying reason for insignificant results of the said specification might be because the variance in each outcome variable is very high for both control and treatment group (see Appendix Figure A4 and Figure A5). Many non-PPG teachers performed better in their CK and PCK tests, and at the same time, there were PPG graduates who performed low on the standardised tests. One should note that this estimation results applied only to teachers at the cutoff and were graduates of the seven teacher colleges in the sample.

4.2. Impact on Student Learning Outcomes

Essentially, our empirical strategy is to compare the outcomes of students taught by PPG graduate teachers to those taught by non-PPG teachers. Since we could not control which grade the teachers were assigned to, we pooled student data across grades. Our ultimate interest is an aggregate impact on students taught by PPG graduate teachers and the non-PPG teachers rather than the grade-specific impact of PPG. We standardised the students' outcome variables using students taught by the non-PPG teachers in the first grade as the base.

We estimated all equations described in Section 3.1.1 using smaller bandwidth and without covariates. For estimation with larger bandwidth, we included covariates. All of the estimations were clustered at the teacher level. Since the intervention is at the teacher level, we also included covariates at the student level that might be correlated with the outcome mean. Classroom and school level data might also be correlated with the outcome. Since our sample teachers are classroom teachers at different schools, each teacher represents one classroom and one school. In the estimation with covariates, we included school quality index as classroom/school level covariate.

Following the RD plot of the covariates, as we expect, the proxy of students' socio-economic background (i.e., their housing quality and parents' education background) did not show discontinuity around the cutoff of 50. However, the RD plot indicates that there might be classroom or school effects, presented by discontinuity of the school quality index around the cutoff. In the estimation, we also considered that there might be differences in how the PPG and non-PPG graduate teachers taught during the COVID-19 pandemic. The RD plot of the said covariate (i.e., the number of days per week the teaching occurs) shows a discontinuity at the cutoff.

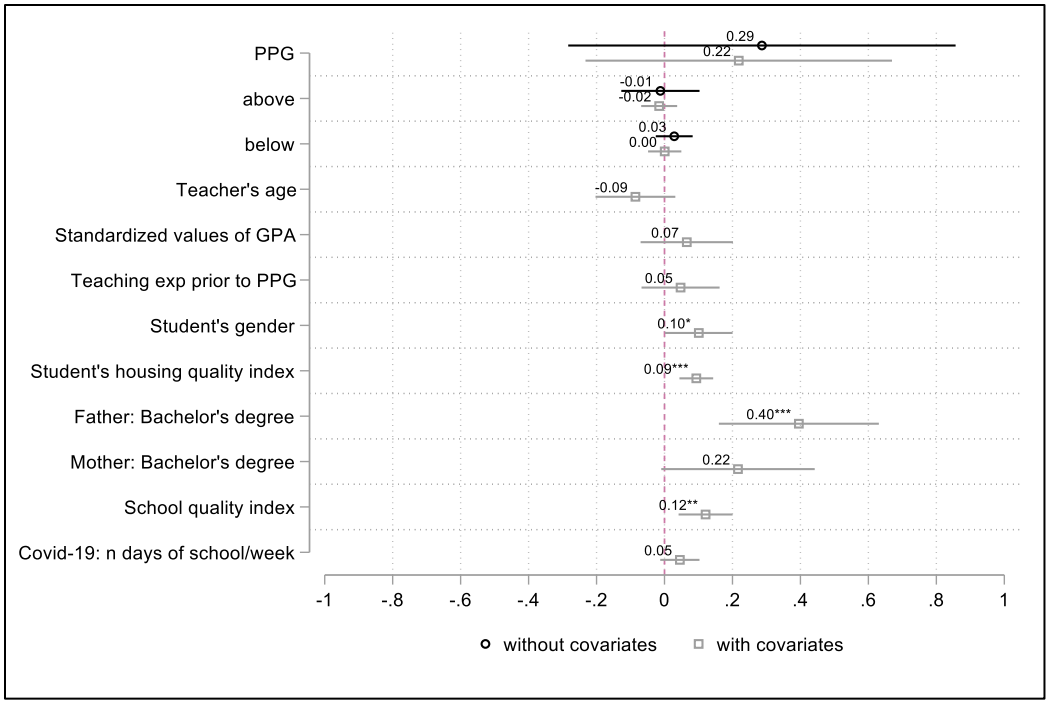

Figure 4. Estimation Results for Students' Standardised Scores in Numeracy

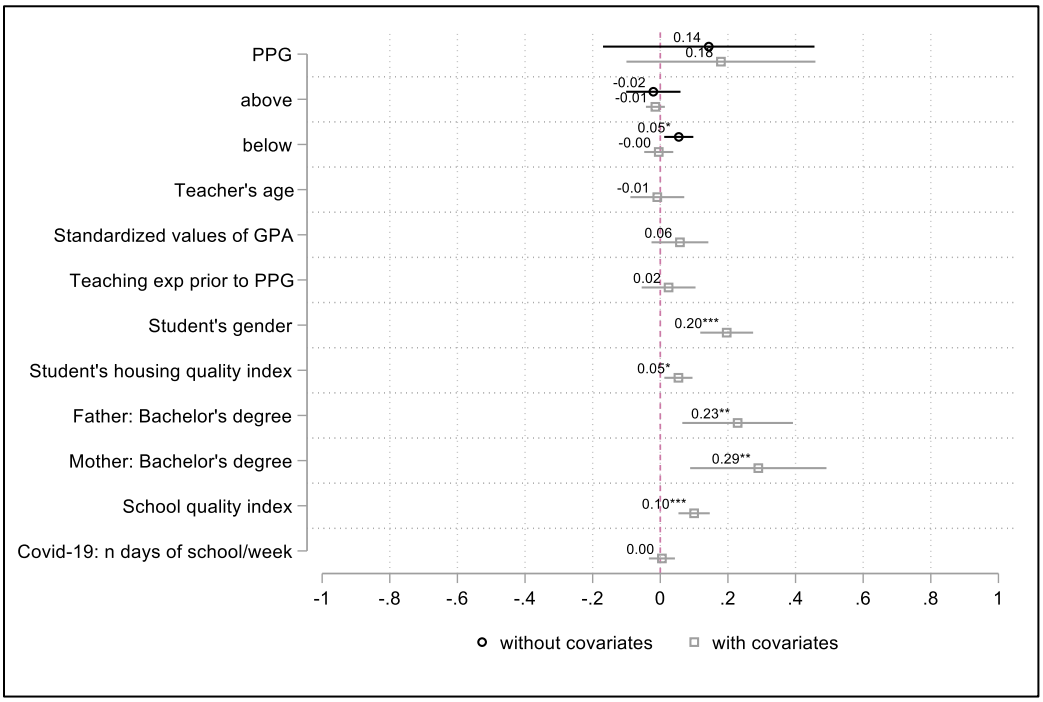

Figure 5. Estimation Results for Students' Standardised Scores in Literacy

Figure 4 and Figure 5 show that although the coefficient of being taught by PPG teachers is positive in the estimation with and without covariates, the results of the said coefficient are insignificant for both students' standardised scores in numeracy and literacy. It suggests that at exactly the point where teachers experienced a jump in the probability of enrolling in PPG, two years later, there was no jump in the numeracy and literacy scores of students taught by these teachers for approximately eight months. Similarly, when we carried out sensitivity analysis using various bandwidth, the results did not change.

Our chosen specification among all estimations is the one with larger bandwidth and covariates. We found no significant impact of being taught by PPG graduates on students' learning outcomes in numeracy and literacy. The results are likely because the distribution of the outcome variables at student levels has large dispersion relative to the impact of training. While much can be explained by quantitative analysis, the results lead to the question of whether training focused on pedagogy differentiates PPG graduate teachers from non-PPG teachers when they teach in schools. Supported by the qualitative data collection, the next subsection provides insight and supporting evidence on the structural design of pre-service PPG from the intake, curricula, to graduation criteria.

4.3. Degree to the Effectiveness of Teacher Training

Various forms of teacher preparation have been implemented globally to ensure teacher quality. Generally, initial teacher education begins with four years of education at a teacher college, followed by an extension programme primarily aimed at providing teacher candidates with actual teaching experience prior to graduation or placement. A teacher preparation programme, at its core, adheres to the same principle. While it might come in different forms, this principle revolves around these specific aspects: programme vision, admission selection, programme curriculum (which includes field experience, clinical supervision, reflective practice, action research), and graduation attributes.

When comparing pre-service PPG features to the global practice of teacher preparation programmes, the pre-service PPG has features that apply these principles differently. Based on the qualitative data, we analyse the extent to which the principles are embedded in PPG features. The analysis consists of several aspects: the vision of the programme, intake, curricula, graduation criteria, and locus of control.

Vision of the Programme

According to its guidelines, PPG's vision is to produce teacher candidates who fully master teacher competencies in accordance with the national education standards, allowing them to obtain professional teacher certificates. In general, our respondents, the teacher candidates, shared a similar understanding of PPG's vision to produce professional and quality teachers. The respondents' knowledge of the PPG's vision is broadly divided into four categories: (1) PPG is a professional education programme for novice teachers following undergraduate teacher education; (2) PPG aims to improve teachers' professionalism and competence to increase student learning achievement; (3) PPG emphasises teaching practice to train prospective teachers before entering schools; and (4) PPG grants certificates to teachers as a guarantee to the public that they are qualified professionals.

We contrasted the shared vision of the teacher candidates with their motivation for pursuing PPG. Based on a short survey in the first year, we found that obtaining teaching certificates is the primary motivation for participating in PPG (see Appendix Figure A4), while improving teaching competency is their second motivation. The motivation for obtaining certification implies that the candidates' personal

goal is to gain administrative advantages for their civil service career trajectory and to receive potential future teacher allowance.

Improvement of Intake

The main feature of PPG and what differentiates it from the initial teacher education (i.e., bachelor's degree in education) in Indonesia is the selection of candidates upon entry. The administrative requirements to pass the next selection process include a minimum GPA of 3.0 out of 4.0 from an undergraduate degree in an accredited university and less than five years of teaching experience. If applicants met the administrative requirements, they would be contacted to undertake an online computer-based test at the local LPTKs. The online admission test is a centralised and standardised exam administered by the MoEC. The online test includes academic aptitude, English skills, and pedagogical knowledge. The online admission test score is a weighted average of the three scores. The weight for academic aptitude, English skills, and pedagogical knowledge test scores are 50 percent, 20 percent, and 30 percent, respectively. Applicants can continue to the next phase of selection, the interview, if they score above 50 on the online admission test. In the interview phase, applicants will be assessed using instruments consisting of fifteen indicators. The interview instrument aims to gauge applicants' knowledge about teacher regulation and teacher competencies; motivation for pursuing a career as a teacher; personality and talent; and attitude as well as appearance. The online admission test score and interview score are weighted and added to generate the final score. The weight for online admission test score is 70 percent, while the interview score is 30 percent.

The cutoff score of 50 to pass the online admission is determined independently by the MoEC.

> *"The selection component is a competency test (i.e., online admission test) consisting of 30 multiple-answer questions, …. . Currently, the passing standard is 50 for the competency test. This is adjusted to the calculation of the MoEC (was Ministry of Research, Technology and Higher Education) that competency test scores increase by 16–20 points in a year. Hence, at the end of PPG, PPG graduates achieve a minimum score of 70."* (Programme Organiser, Female)

The programme organiser set the cutoff score at 50 based on the assumption that when teacher candidates graduate, they will achieve a score of 70 in their exit exams. A score of 70 is said to be equivalent to the Teachers Competency Test (*Uji Kompetensi Guru* or UKG). A study by Yusrina et al. (2022) shows that the online admission test can predict teacher candidates' exit exam at the end of PPG.

Looking at a teacher preparation model explained by Musset (2010), the pre-service PPG is a consecutive model. In particular, the pre-service PPG for PSTE follows the said model because it allows candidates who have obtained a degree in PSTE and from the relevant disciplines. A distinctive feature of the consecutive model is that it relies on a stronger foundation of previously acquired subject matter knowledge, as it requires candidates to hold a relevant degree (Zuzovsky and Donitsa-Schmidt, 2017). Such information reveals that all the accepted candidates are considered to have a strong background in the relevant CK. However, the selection process did not specifically assess relevant CK, such as literacy and numeracy, to screen candidates with the expected strong background.

In addition to the entry requirements, based on our analysis of the administrative data in the selection process, 39 out of 4,339 applicants did not major in PSTE, but rather a relevant discipline—mostly took PSTE in religious education (madrasa). Although the design encourages non-PTSE to apply to PPG, the

1,291 applicants admitted to the program are all undergraduates from PSTE. In the next section, we mention the implication of having all teacher candidates entering PPG come only from PSTE.

Improvement in Curricula

The pre-service PPG curriculum was structured into one semester of in-class sessions and one semester of school placement and action research. According to the PPG curriculum and guidelines developed by the MoEC, during the first semester, the teacher candidates in the PSTE learn in small groups how to prepare lesson plans, develop students' worksheets, use appropriate teaching aids, and develop the corresponding assessments for five subjects taught in Grades 1 to 6 primary school (math, language, natural science, social science, and civic education). The candidates are also required to perform peer teaching on selected lessons during this period.

From our interviews with teacher candidates and teacher educators, a particular theme emerges in regard to their perception of the in-class session. The in-class session is perceived to be monotonous and repetitive to what teacher candidates received during their undergraduate PSTE. During this first semester, the respondents associated the said session with only practicing the use of technology in the classroom and developing lesson plans, which also included determining and developing the appropriate teaching aids.

As briefly described in the programme intake, all accepted applicants are PSTE undergraduates. They felt that the components in the in-class sessions had been addressed over the course of four years of undergraduate study, suggesting that the curricula of the in-class sessions are not optimal for candidates with a PSTE background.

The teacher candidates suggested the expected curriculum that may affect their teaching quality, which consists of skills or knowledge that prepare them to accommodate the needs of a diverse student population. Similarly, the mentor teachers, who are responsible for supervising teacher candidates during an internship at school, called attention to the lack of teacher candidates' acknowledgment of diversity of students in the classroom. The mentor teachers recommended such knowledge to be in the in-class session before the internship.

For approximately six months, teacher candidates do their internship at a school partner in the second semester. The candidates are placed at partner schools in groups of five, supervised by a teacher educator and a mentor teacher. For each teacher candidate, PPG requires a minimum of eight classroom observation visits by both teacher educator and mentor teacher.

The PPG guidelines provide instruments on aspects that must be scored to ensure teacher candidates are assessed uniformly during observations. However, the instruments are not complemented by relevant rubrics that guide mentor teachers in scoring. Interviews with teacher candidates revealed that they were only given scores and were less likely to receive feedback comments.

According to a university organiser, besides getting a mentored internship, candidates also have the opportunity to conduct research, reflect on their teaching practice, discuss with their mentor teacher, as well as experience non-teaching activities, such as school management and extracurricular activities.

> "*During the internship, ideally, we have clinical supervision. Clinical supervision is when student and mentor teacher can learn together about the best practice throughout the internship when dealing with existing or potential issues that may emerge as they become a teacher. Teacher candidates must at least*

> *conduct one session of clinical supervision throughout the semester where they are being observed by their mentor teacher followed by a reflection session."* (University Organiser, Male, 14 December 2018)

The PPG curriculum imposes that during school-based activities, the clinical supervision and reflection of instructional practice should be used. Research has identified a combination of such approaches as conducive to developing PCK (Darling-Hammond, 2006; De Jong and Van Driel, 2004). If such a curriculum is conducted properly, it will be embedded and carried through by PPG graduate teachers in their in-service teaching.

Improvement of Graduation Criteria

The PPG's graduation criteria differ significantly from an undergraduate degree in education. PPG includes performance assessment, which is deemed to contribute to programme effectiveness (Valli and Renert-Ariev, 2002). The PPG's exit exams are centralised, and candidates can earn a teaching certificate when they graduate. To graduate, teacher candidates must take two exit exams administered by the MoEC:

- The teaching practice exam, *Uji Kinerja* (UKin), a performance test, aims to assess a teacher candidate's ability to develop lesson plans and select relevant teaching aids, implement the said lesson plan, and evaluate learning. A candidates' performance in this exam is assessed by two evaluators, a teacher educator from the LPTK and a mentor teacher from the partner school. Neither of the evaluators is the candidate's teacher educator or mentor teacher. The instruments for this practice exam are developed centrally by the MoEC. Prior to the exams, the MoEC provides training for the evaluators to ensure a uniform perception of the assessment indicators.
- The knowledge exam, *Uji Pengetahuan* (UP), is a written test that includes numeracy, Indonesian language, science, civic education, methods in developing assessment instruments, and differentiated instruction.

Although the exit exams are also teaching licensure and thus considered high-stake, all candidates eventually graduated and were awarded teacher certification. The candidates were allowed to revisit the test if they failed the first attempt. In the 2018 pre-service PPG PSTE cohort that we observed, 94.27% of the 1,291 candidates succeeded on the first attempt.

The highlight from the graduation criteria is the fact that the teacher candidates did not prepare to have a CK test in the exit exams. In their argument, the one-year training was more about the craft of teaching and having the experiences to smooth their transition.

> *"My complaint with the PPG is that we had to pass the knowledge exam (UP), while the programme was more about the process (of teaching). So many of my peers focused on the UP. For example, during peer teaching, my peers would rather study for UP. They thought peer teaching did not affect [their chances of] passing, while the UP result did."* (Teacher T, Female, 10 December 2018)

Some LPTKs addressed the need of their candidates by setting up additional CK classes in to drill candidates so they can pass the knowledge exam. In fact, the CK classes did not intend to provide candidates with thorough understanding of the contents for future teaching references.

Locus of Control

The regulatory policy about teacher preparation programme was about who controls the design and is in charge of teacher education delivery. The PPG curriculum and guidelines are designed to enable LPTKs

(which were selected to manage PPG) to make adjustments based on the institution's capabilities. The MoEC developed the design with general guidance to accommodate variations in the quality of LPTKs throughout Indonesia. Despite the LPTKs being selected, they differ in their capacity to shape and ensure the quality of PPG graduate teachers. Even so, the selection of the teacher colleges is subject to observable characteristics, such as accreditation status, staff, facilities, and less on how learning in the college takes place or the quality of its graduates.

The locus of authority matters for quality (Tatto et al., 2012). Once applicants are accepted, LPTKs are responsible for quality issues— which refers to criteria they did not create. Accordingly, regardless of how the MoEC controls the quality of the PPG entrants, it will be up to the LPTK to determine the outcome.

5. Discussion and Conclusion

The reform in pre-service training shows traits that address criticism of the quality of bachelor's degree in education. The selection mechanism of prospective teachers on PPG allows for identification of impact of the programme on teacher effectiveness. The cutoff point used in the selection provides identification of prospective teachers accepted into PPG and the comparison group (i.e., the non-PPG teachers). We measured teacher effectiveness using teachers' standardised scores in CK and PCK tests, both in numeracy and literacy. The tests were administered to both groups two years after they graduated and already had teaching positions. At the same time, we also measured the learning outcomes of students taught by PPG and non-PPG graduate teachers for approximately eight months after the teaching started. Accordingly, we estimated the impact of PPG on the standardised students' scores in numeracy and literacy.

Since we compare PPG graduate teachers to their non-PPG counterparts, we hypothesise that PPG improves competency of prospective teachers. In the sampling, our PPG teachers earned their undergraduate degrees at the same teacher college as their non-PPG counterparts. We expect teachers in the two groups to likely have similar characteristics before joining PPG. Using fuzzy RDD, our preferred specification is the estimation with larger bandwidth and covariates. As we use larger bandwidth, we include more samples on both sides of the cutoff score. Accordingly, we include covariates to control for other factors that might create discontinuity around the cutoff. The estimation of student scores is clustered at the teacher level.

We found that local estimators for numeracy content and PCK estimates were negative, but student scores were positive. As for literacy content knowledge, it was negative, the PCK was positive, and the student scores were also positive. The result also shows that the variance in the distribution of teacher outcomes in numeracy is higher relative to literacy content. Such a pattern is similar to the previous literature (Goldhaber et al., 2013; Kane and Staiger, 2012) that the variance in the distribution of teacher quality in numeracy is likely to be higher than in literacy.

Overall, the estimation with larger bandwidth and covariates for the teachers' knowledge and student learning outcomes show no effect on PPG graduate teachers. Potentially, the effect is missing due to the low power results. However, based on the kernel density plot (Appendix), we see that the variance in each of the outcome variables is relatively high for both the treated and control group, and there exists massive overlap, suggesting that the likelihood of inclusion and exclusion errors is high. The inclusion

error means that the selection might have included candidates when they should not have. The exclusion error refers to candidates who were not accepted into the programme when they should have. We argue that PPG as "a game" to create a proxy for quality teachers is almost worthless since both errors can be found in the PPG and non-PPG teachers. Our results contrast with a study by Goldhaber and Startz (2017), in which the authors argue that a preparation programme that focuses on the tail of the distribution, particularly the high-performing one, is likely to change teacher quality upwards.

The likelihood of inclusion and exclusion errors at the selection can be referred to the selection mechanism. Being a consecutive model, the PPG selection of its candidates should rely on a stronger foundation of previously acquired subject matter knowledge (Zuzovsky and Donitsa-Schmidt, 2017). In our case, the candidates for PPG in PSTE major should have been selected based on their CK in numeracy and literacy. A previous study on pre-service PPG reveals that the online admission test, used to screen the candidates for the 2018 cohorts, is not associated with student learning outcomes (Yusrina, et al., 2022). Thus, recommend that future PPG candidates majoring in PSTE, and to address the inclusion and exclusion errors at selection, be selected based on their CK in numeracy and literacy.

The plausible explanation of why PPG has no impact on teacher competency might be attributed to the programme's design. By design, the PPG curriculum has included several features of an effective teacher preparation programme. These features have been recognised, to a certain extent, to be critical for teacher preparation programme as they are identified as predictors of teaching performance (Reagan et al., 2021; Popova et al., 2019; Tatto, 2015; Valli and Renert-Ariev, 2002). The PPG curriculum has included longer school-based activities, a clinical supervision approach, and instructional practice reflection. Seeing as such practices have been incorporated, the programme design appears capable of producing higher-quality teachers when in fact, it is not. This is an isomorphic mimicry phenomenon (Andrews et al., 2017).

The MoEC provides general guidelines for implementing the PPG, such that teacher colleges can interpret them and implement the curriculum based on each college's ability. We acknowledge that there are variations in how LPTKs implement those best practices. However, we suggest minimising the variation by improving the programme design rather than pointing at each LPTK's wrongdoing.

A recent international study of pre-service teacher preparation shows a correlation between central control in the provision of the programme and teachers' outcomes in math content and PCK (Tatto et al., 2012). Suggesting that the extent to which the best practices in the PPG training are effective requires more control from the MoEC. It might include improving assessment instruments and providing relevant rubrics, specific guidance, training, and monitoring to the university organiser and partner schools for conducting practices such as clinical supervision and reflection session.

Based on the discussion, our study presents two significances. First, our findings show that if there is a way to conduct initial teacher training that improves teacher practices that lead to higher student outcomes, PPG was not it. Second, the completion of PPG is high stakes. If replicating best practices embedded in the training becomes merely a checklist without greater accountability, we cannot expect it to build teachers' true capability. We suggest the following ways to improve the effectiveness of pre-service PPG for PSTE: (1) Conduct a more effective screening of teacher candidates using content knowledge tests in numeracy and literacy; (2) Provide relevant training content to ensure a closer link between university courses and practical classroom teaching in schools, such as in-class session that

focuses on meeting the needs of a diverse student population; (3) Strengthen the MoEC's control in delegating effective practices in the programme, such as a standardised guideline for encouraging reflective sessions and providing rubrics for the assessment instruments to encourage meaningful feedback in a clinical supervision session.

References

Andrews, M., Pritchett, L. and Woolcock, M. (201)7. *Building State Capability: Evidence, Analysis, Action*. Oxford: Oxford University Press, p.288.

Best, A., Tournier, B. and Chimier, C. (2018). *Topical Questions on Teacher Management*. [online] International Institute for Educational Planning UNESCO. Available at: http://www.iiep.unesco.org/sites/default/files/topical_questions_on_teacher_management_english.pdf [Accessed 5 December 2021].

Blömeke, S., Olsen, R. and Suhl, U. (2016). 'Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality' in Nilsen, T. and Gustafsson, J. (eds). *Teacher Quality, Instructional Quality and Student Outcomes. IEA Research for Education, vol 2*. Springer. pp.21-50.

Boyd, D., Grossman, P., Lankford, H., Loeb, S. and Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Educational Evaluation and Policy Analysis*, 31(4), pp.416-440.

Calonico, S., Cattaneo, M., Farrell, M. and Titiunik, R. (2017). Rdrobust: Software for Regression-Discontinuity Designs. *The Stata Journal: Promoting communications on statistics and Stata*, 17(2), pp.372-404.

Canales, A. and Maldonado, L. (2018). Teacher Quality and Student Achievement in Chile: Linking Teachers' Contribution and Observable Characteristics. *International Journal of Educational Development*, 60, pp.33-50.

Cattaneo, M., Jansson, M. and Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal: Promoting communications on statistics and Stata*, 18(1), pp.234-261.

Chetty, R., Friedman, J. and Rockoff, J. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), pp.2633-2679.

Clotfelter, C., Ladd, H. and Vigdor, J. (2007). Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. *Economics of Education Review*, 26(6), pp.673-682.

Cochran-Smith, M. and Maria Villegas, A. (2015). Studying Teacher Preparation: The Questions that Drive Research. *European Educational Research Journal*, 14(5), pp.379-394.

D'Agostino, J. and Powers, S. (2009). Predicting Teacher Performance with Test Scores and Grade Point Average: A Meta-Analysis. *American Educational Research Journal*, 46(1), pp.146-182.

Darling-Hammond, L. (2006). Assessing Teacher Education: The Usefulness of Multiple Measures for Assessing Program Outcomes. *Journal of Teacher Education*, 57(2), pp.120-138

Darling-Hammond, L. (2016). Research on Teaching and Teacher Education and Its Influences on Policy and Practice. *Educational Researcher*, 45(2), pp.83-91.

Darling-Hammond, L. (2017). Teacher Education around the World: What can We Learn from International Practice?. *European Journal of Teacher Education*, 40(3), pp.291-309.

De Jong, O. and Van Driel, J. (2004). Exploring the Development of Student Teachers' PCK of the Multiple Meanings of Chemistry Topics. *International Journal of Science and Mathematics Education*, 2, pp.477-491.

De Ree, J., Muralidharan, K., Pradhan, M. and Rogers, H. (2018). Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia. *The Quarterly Journal of Economics*, 133(2), pp.993-1039.

Döhrmann, M., Kaiser, G. and Blömeke, S. (2012). The Conceptualisation of Mathematics Competencies in the International Teacher Education Study TEDS-M. *ZDM Mathematics Education*, 44, pp.325-340.

ESSPIN. (2008). *An Assessment of the Development Needs of Teachers in Nigeria - Kwara State Case Study*. Report No. KW 301. [online] Education Sector Support Programme in Nigeria. Available at: http://www.esspin.org/resources/reports/kwara

Frölich, M. and Huber, M. (2018). Regression Discontinuity Design with Covariates. *Journal of Business & Economic Statistics*, 37(4), pp.736-748.

Giallo, R. and Little, E. (2003). Classroom Behaviour Problems: The Relationship between Preparedness, Classroom Experiences, and Self-Efficacy in Graduate and Student Teachers. *Australian Journal of Educational & Developmental Psychology*, 3, pp.21-34.

Goldhaber, D. and Startz, R. (2017). On the Distribution of Worker Productivity: The Case of Teacher Effectiveness and Student Achievement. *Statistics and Public Policy*, 4(1), pp.1-12.

Goldhaber, D., Liddle, S. and Theobald, R. (2013). The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement. *Economics of Education Review*, 34, pp.29-44.

Harris, D. and Sass, T. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics*, 95(7-8), pp.798-812.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, 79(3), pp.933-959.

Ingvarson, L. and Rowley, G. (2017). Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. *Educational Researcher*, 46(4), pp.177-193.

Jacob, B. and Lefgren, L. (2004). Remedial Education and Student Achievement: A Regression-Discontinuity Analysis. *Review of Economics and Statistics*, 86(1), pp.226-244.

Kane, T., and Staiger, D. (2012). *Gathering Feedback for Teaching Combining High-Quality Observations with Student Surveys and Achievement Gains*. Bill and Melinda Gates Foundation. [online] EdTech Hub. Available at: https://files.eric.ed.gov/fulltext/ED540960.pdf

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S. and Baumert, J. (2013). Teachers' Content Knowledge and Pedagogical Content Knowledge: The Role of Structural Differences in Teacher Education. *Journal of Teacher Education*, 64(1), pp.90-106.

Mincu, M. (2015). Teacher quality and school improvement: what is the role of research?. *Oxford Review of Education*, 41(2), pp.253-269.

Musset, P. (2010). *Initial Teacher Education and Continuing Training Policies in a Comparative Perspective: Current Practices in OECD Countries and a Literature Review on Potential Effects*. OECD Education Working Papers 48. [online] OECD Publishing. Available at: https://www.oecd-ilibrary.org/education/initial-teacher-education-and-continuing-training-policies-in-a-comparative-perspective_5kmbphh7s47h-en

Popova, A., Evans, D., Breeding, M. and Arancibia, V. (2019). *Teacher Professional Development around the World: The Gap between Evidence and Practice*. Working Paper 517. [online] Washington, D.C.: Center for Global Development. Available at: https://www.cgdev.org/publication/teacher-professional-development-around-world-gap-between-evidence-and-practice

Reagan. E.M., Ahn, J., Roegman, R., and Vernikoff, L. (2021). What Makes Teacher Preparation Legitimate? An Analysis of Teacher Residency Websites. *Action in Teacher Education*, 43(2), pp.144-159.

Rivkin, S., Hanushek, E. and Kain, J. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), pp.417-458.

Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, 94(2), pp.247-252.

Shulman, L. (1986). Those Who Understand: A Conception of Teacher Knowledge. *American Educator*, 10(1).

Tatto, M. (2015). The Role of Research in the Policy and Practice of Quality Teacher Education: An International Review. *Oxford Review of Education*, 41(2), pp.171-201.

Tatto, M., Peck, R., Schwille, J., Bankov, K., Senk, S., Rodriguez, M., Ingvarson, L., Reckase, M. and Rowley, G. (2012). *Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Valli, L. and Renert-Ariev, P. (2002). New standards and assessments? Curriculum transformation in teacher education. *Curriculum Studies*, 34 (2), 201-225.

Westbrook, J., Durrani, N., Brown, R., Orr, D., Pryor, J., Boddy, J. and Salvi, F. (2013). *Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries*. Final Report. [online] DFID Education Rigorous Literature Reviews. Available at: http://10.13140/RG.2.2.27180.77446

World Bank. (2010). *Transforming Indonesia's Teaching Force: Executive Summary*. World Bank. [online] World Bank Group Open Knowledge Repository. Available at: https://openknowledge.worldbank.org/handle/10986/2852

Yusrina, A., Bima, L., Berkhout. E., and Suryadarma, D. (2022). *Selecting Teachers in Indonesia: Predicting Teacher Performance. Using Pre-employment Information*. RISE Working Paper 22/098. [online] RISE Programme. Available at: https://riseprogramme.org/sites/default/files/2022-05/Pre-Service%20Teachers%20Selectivity.pdf

Zuzovsky, R. and Donitsa-Schmidt, S. (2017). Comparing the effectiveness of two models of initial teacher education programmes in Israel: concurrent vs. consecutive. *European Journal of Teacher Education*, 40(3), pp.413-431.

Appendix

Table A1. Tracing Results of Applicants of Pre-Service PPG Programme 2018

Teacher Colleges	Targeted sample from administrative data in 2007		Tracing results using phone survey in 2020		Field visit results in 2021	
	Not accepted	Accepted	Not accepted	Accepted	Not accepted	Accepted
Universitas Muhammadiyah Malang	42	6	**5**	5	**4**	5
Universitas Muhammadiyah Purwokerto	70	4	**13**	3	**12**	3
Universitas Negeri Jakarta	49	26	**10**	16	**8**	14
Universitas Negeri Surabaya	46	51	**8**	38	**4**	38
Universitas Pendidikan Indonesia	202	47	**35**	23	**34**	17
Universitas Sanata Dharma	34	31	**9**	26	**7**	24
Universitas Sebelas Maret	48	32	**7**	21	**6**	21
Subtotal	491	197	**87**	132	**75**	122
Total	688		219		197	

Notes: In terms of gender, age, and undergraduate GPA (the observable covariates from administrative dataset), there are no significant differences between the 688 and the 197-sample group.

Table A2. Reasons of Choosing School

Reasons of choosing the current school	Teachers graduates from PPG		Teachers - non PPG	
	Freq	Percent	Freq	Percent
In a close distance	51	0.42	30	0.41
School achievement	4	0.03	3	0.04
Availability of vacancy	50	0.41	32	0.43
Others	16	0.13	9	0.12
	121		74	

Table A3. Descriptive Statistics on the Observable Covariates for All Sample and within Selected Bandwidth

Variable	Teachers graduate from PPG					Non-PPG teachers				
	n	Mean	sd	[95% Conf. Interval]		n	Mean	sd	[95% Conf. Interval]	
Full sample										
Grade 1	11					2				
Grade 2	10					3				
Grade 3	15					15				
Grade 4	23					11				
Grade 5	32					20				
Grade 6	23					16				
Civil servant	101					33				
Non-Civil servant	21					42				
Bandwidth -5<c<5										
Gender	46	0.78	0.42	0.66	0.91	43	0.67	0.47	0.53	0.82
Age	46	24.57	1.31	24.18	24.95	43	24.72	1.39	24.29	25.15
Undergraduate GPA	46	3.61	0.18	3.55	3.66	43	3.52	0.14	3.48	3.56
Years of teaching experience prior to PPG	45	2.07	1.3	1.67	2.46	42	2.62	1.17	2.25	2.98
School quality index										
Grade 1	5					2				
Grade 2	4					2				
Grade 3	6					7				
Grade 4	8					6				
Grade 5	14					12				
Grade 6	5					10				
Civil servant	33					23				
Non-Civil servant	13					20				

Table A4. List of Qualitative Respondents and Methodology

Year of Data Collection	Themes	Respondent										
		Teachers (PPG students)	Teachers (non-PPG)	MoRTHE	PPG Lecturers	PPG Organisers	Mentor Teachers	Partnered School Principal	Students	Principal	Parents	Classroom observation
2018	PPG Design			x	x	x						
	PPG Implementation	x		x	x	x	x	x				□
	Factors Influencing Teaching performance	x			x	x	x	x				
	PPG Added Value	x			x	x	x	x				
2020/2021[2]	Teaching Practices	x	x						o[3], x	x	x	□
	Factors Influencing Teaching Practices	x	x							x		
	Perception on PPG Teachers	x							x	x	x	
	PPG Added Value	x	x							x		
N (data collection)		**31**	**9**	**1**	**12**	**17**	**15**	**13**	**5, 59**	**24**	**73**	**17**

Notes: x: in-depth interview
o: focused group discussion
□: student classroom observation

Due to the COVID-19 pandemic, we made several adjustments to the data collection. In 2018, we had 16 teacher respondents from the treatment group whom we planned to re-interview in 2020. During our data collection in March 2020, however, there was a school closure policy because of the COVID-19 pandemic. As a result, we were only able to visit nine of the fifteen schools. Eight teachers from the treatment group and one from the control group were interviewed. Out of nine FGDs planned, we managed to conduct five FGDs with students in 2020.

[2] In 2021, we added information on Schooling from Home practices.

[3] Due to the Pandemic, FGD with students were changed into interview with four selected students

We resumed the data collection in March–April 2021 with several changes and adjustments, such as the number of respondents and data collection methods. We decided to add more respondents to the control group to collect more information about factors influencing teaching quality. In total, we have 15 teachers' respondents (seven from the treatment group and eight from the control group). Due to school closure in 2021, it was not possible to perform FGD with kids because it was not possible to assemble them at school. We changed the FGDs into interviews with students at their homes. To identify how PPG and non-PPG teachers modified their teaching in response to the school closure, we also adjusted the instruments and asked a few questions about their experiences as students during the pandemic. Due to the fact that face-to-face learning allowed students to interact directly with their teacher, they tended to favour their former teacher over our teachers' respondents.

RD Manipulation test using local polynomial density estimation.

c = 0.000	Left of c	Right of c
Number of obs	66	131
Eff. Number of obs	42	43
Order est. (p)	2	2
Order bias (q)	3	3
BW est. (h)	5.187	4.538

Number of obs	= 197
Model	= unrestricted
BW method	= comb
Kernel	= uniform
VCE method	= jackknife

Running variable: c_oat.

Method	T	P>|T|
Robust	-0.3081	0.7580

P-values of binomial tests. (H0: prob = .5)

Window Length	<c	>=c	P>|T|
1.429 + 1.429	10	10	1.0000
1.846 + 1.774	15	13	0.8506
2.264 + 2.120	23	19	0.6440
2.681 + 2.465	25	19	0.4514
3.099 + 2.810	28	23	0.5758
3.516 + 3.156	30	28	0.8957
3.934 + 3.501	34	29	0.6147
4.352 + 3.847	37	33	0.7202
4.769 + 4.192	38	38	1.0000
5.187 + 4.538	42	43	1.0000

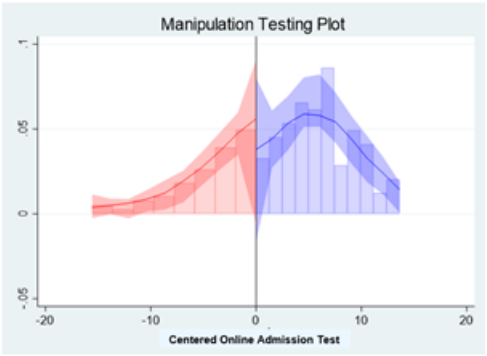

Figure A1. Manipulation Testing Plot

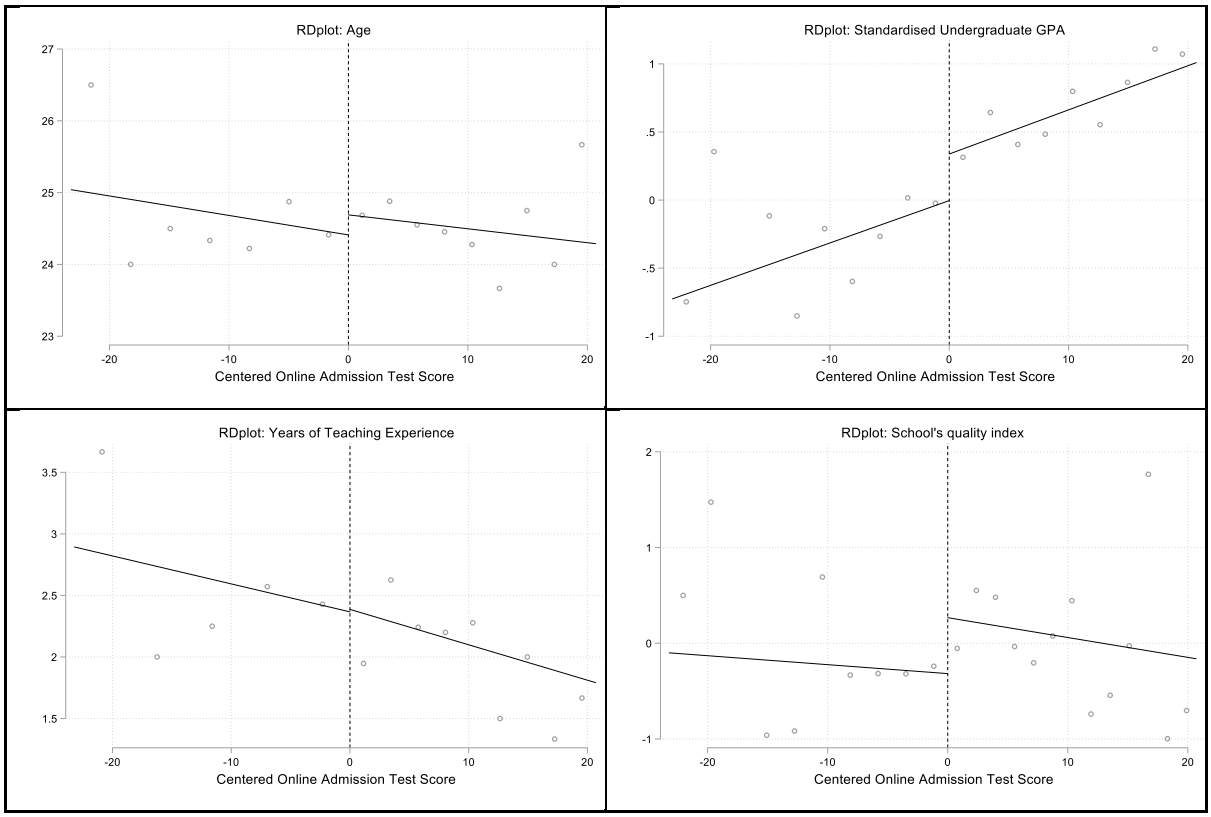

Figure A2. Falsification Graphs on Teacher Covariates

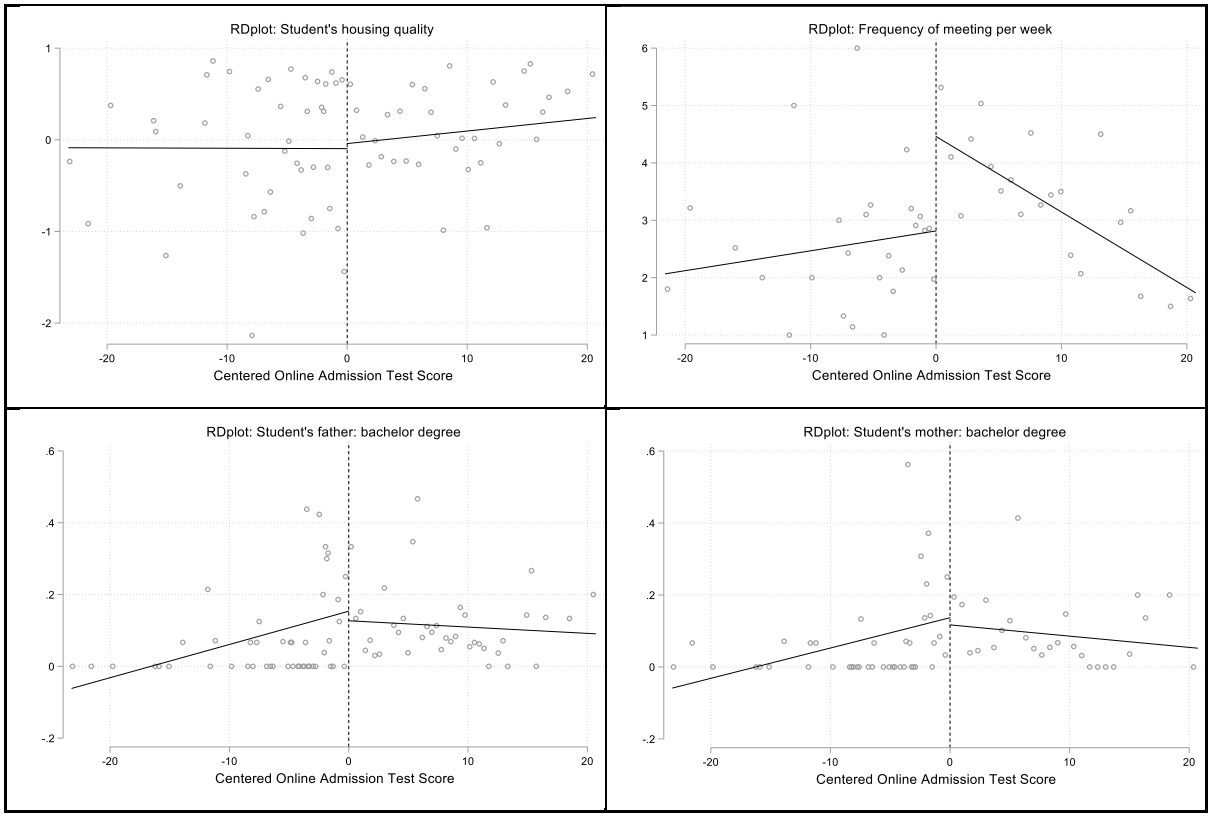

Figure A3. Falsification Graphs on Student Covariates

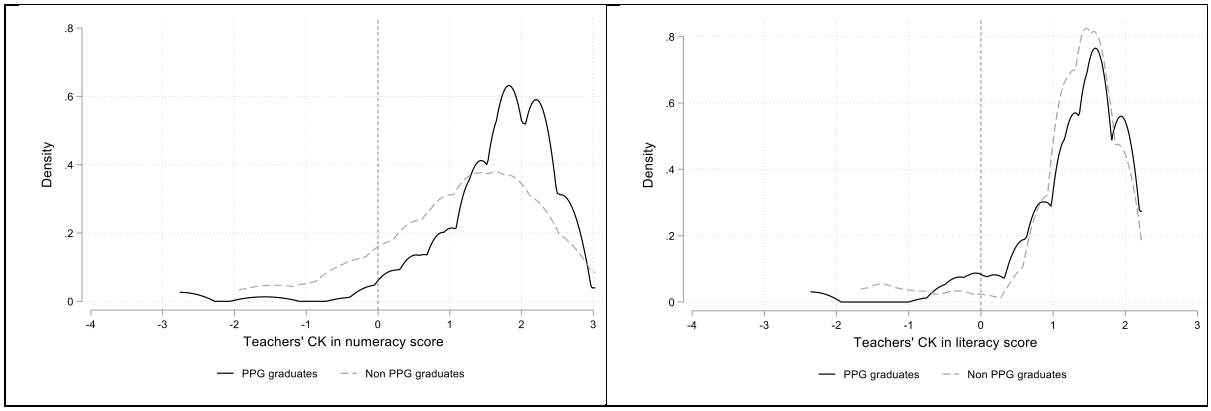

Figure A4. Kernel Density of Teacher's CK score in Numeracy and Literacy

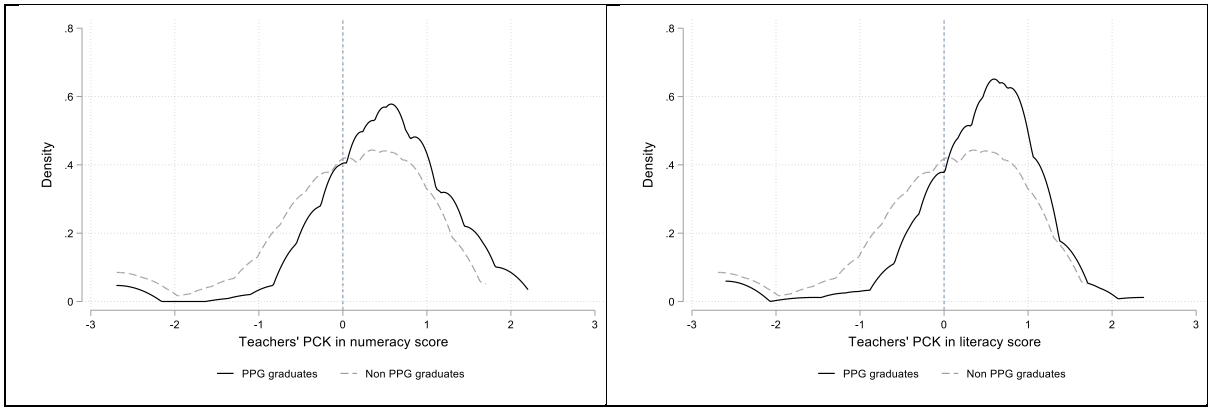

Figure A5. Kernel Density of Teacher's PCK score in Numeracy and Literacy

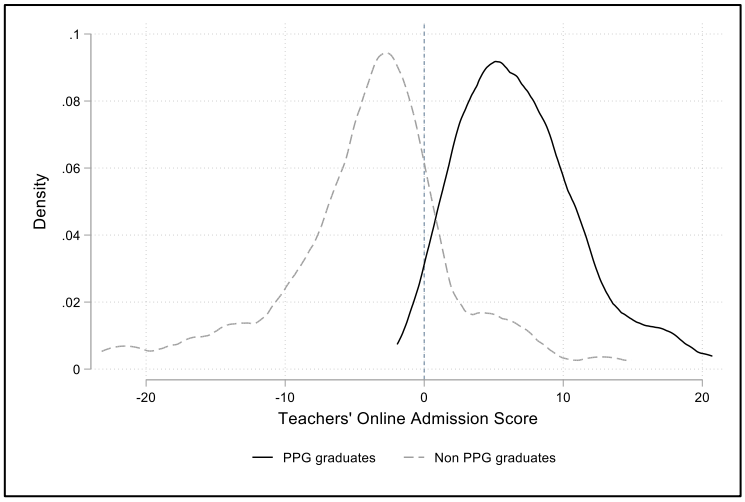

Figure A5. Kernel Density of Online Admission Score