Selecting Teachers in Indonesia: Predicting Teacher Performance using Pre-Employment Information

Prepared by Asri Yusrina, Luhur Bima, Emilie Berkhout, Daniel Suryadarma

Reviewed by Shintia Revina, Sudarno Sumarto

RISE CRT Indonesia

Submitted December 2020

Abstract

A wave of teacher retirement in Indonesia provides an opportunity to replace them with better-performing teachers. We study whether teacher candidates' screening tests into *Pendidikan Profesi Guru* (PPG) or Teacher Professional Education, a postgraduate education programme in Indonesia, can predict their performance at the end of the programme and in an actual classroom situation at the beginning of their teaching career. Using administrative data of 1,291 primary school teacher candidates, we find that admission criteria, including undergraduate grade point average (GPA) and online admission tests and interview scores, can predict a candidate's performance on their knowledge and teaching practice exams at the end of their education programme. A one standard deviation higher online admission test score is associated with a 0.30 standard deviation higher score in the knowledge examination. Teacher candidates with a one standard deviation higher interview score perform 0.07 standard deviation better on the teaching practice examination. For teacher candidates with one standard deviation higher undergraduate GPA, their knowledge examination performance is 0.15–0.17 standard deviation higher on average, and their teaching practice exam score is 0.06-0.07 standard deviation higher on average. We then estimate the predictive ability of the admission criteria on student learning outcomes in numeracy and literacy, which uses 1,530 randomly sampled students taught by 114 teacher candidates. We find no evidence that the selection criteria predicted student learning in a meaningful way. Our results contribute to a nascent body of research on the selection of teachers using ex-ante criteria to identify effective teachers in developing countries.

Keywords: teacher selection, ex-ante, teacher education, admission criteria, exit exams, student learning outcomes

Table of Contents

Tables

Figures

Abbreviations

CERDAS	Classroom Observation Tool for Assessing the Dimensions of Teaching Practices
CERMAT	Comprehensive Reading and Numeracy Assessment Tools
DAPODIK (*Data Pokok Pendidikan*)	Ministry of Education and Culture's Education Data Centre
GPA	grade point average
GTK (*Guru dan Tenaga Kependidikan*)	Teachers and Education Personnel
IRT	Item Response Theory
LPTK *(Lembaga Pendidikan Tenaga Kependidikan)*	teacher education institutions
MoEC	Ministry of Education and Culture
PCA	principal component analysis
PDSPK (*Pusat Data Dan Statistik Pendidikan Kebudayaan*)	Data and Statistics Centre of Education and Culture
PIRL	Progress in International Reading Literacy Study
PPG *(Pendidikan Profesi Guru)*	Teacher Professional Education
PPL (*Praktik Pengalaman Lapangan*)	internship
SKB *(Seleksi Kompetensi Bidang)*	Field Competency Selection
TIMSS	Third International Mathematics and Science Study

1. Introduction

Teachers are an essential element for the success of an education system in improving student learning outcomes. Studies found large variation in teacher quality, suggesting that selecting the best teachers could be a strategy to improve learning outcomes. Yet, selecting effective teachers has proven to be complicated since standard selection criteria like their educational attainment and experience cannot explain their quality (Bau and Das, 2020; Rivkin et al., 2005). Therefore, a key policy-relevant issue is whether education stakeholders can use other screening criteria to predict teachers' future teaching effectiveness.

Selecting effective teachers is especially pressing in developing countries. First, education quality in developing countries is significantly lower. A rapid catch-up requires effective teachers. Since the average teacher effectiveness in developing countries is low, recruiting more effective teachers is crucial. Second, hiring ineffective teachers would drain an already limited education budget. Sustained underperformance could make policymakers divert resources away from education to sectors that appear to be more effective. Finally, low education quality significantly increases school dropout in developing countries (Hanushek et al., 2008) and produces little learning (World Bank, 2019).

Ideally, teacher selection takes place before entering the workforce. This is important since teacher colleges are responsible for providing the pool of candidates to be recruited (Davies et al., 2016; Ackley et al., 2007). Teacher colleges that carry out early selection of their candidates send a signal to schools about the quality of their teacher graduates (Casey and Childs, 2011). Moreover, Hwa and Pritchett (2021) argued that selecting teacher candidates at their education stage could strengthen the perception that teaching is a quality-oriented job and early selection reduces the cost of nurturing teachers' performance throughout their career. Selection into teacher education programmes is also beneficial when the recruitment system cannot consistently hire highly qualified aspiring teachers (Huang et al., 2020). Such a selection plays a role in screening out unqualified teacher candidates when the recruitment system, whether for civil servant or contract teachers, is unable to do so.

In this paper, we study whether selection criteria into a teacher education programme can predict teacher performance in Indonesia. The Government of Indonesia reformed the teacher training system by introducing the teacher professional education programme, *Pendidikan Profesi Guru* (PPG), in 2017. Novice teachers with less than five years of teaching experience are required to enrol in the one-year PPG following a four-year bachelor's degree in education.[1] This programme's admission criteria are expected to select applicants who will complete the programme and become high-quality teachers. The PPG uses undergraduate grade point average (GPA), a set of online admission tests, and an interview as screening measures. The online admission test comprises aptitude, English, and pedagogical knowledge tests. We study the effectiveness of these admission criteria in predicting a candidate's performance on the final PPG examinations and their later performance as teachers in the classroom.

The selection of teachers is especially relevant for the Indonesian context, as the Ministry of Education and Culture (MoEC) data show that the number of teachers entering retirement in 2020 reached 39,094.

---

[1] The one-year PPG programme is also referred as pre-service PPG. It aims to produce teachers competent at planning, conducting, and assessing learning; following up assessment results; guiding and training students; conducting research; and developing professionalism sustainably (Kementerian Riset, Teknologi, dan Pendidikan Tinggi, 2017a). These are the characteristics of a professional teacher whom we expect to be the description of a good teacher.

The number is predicted to increase to almost 90,000 by 2024.[2] The central government assessed the subject knowledge and pedagogical skills of nearly three million teachers with a competency test in 2015 and found that the teachers scored significantly below the passing threshold (Rosser, 2018). Also, many of the retiring teachers had lower national exam scores than cohorts who enrolled more recently (Chang et al., 2014). Indonesia has the opportunity to replace these average quality teachers with better ones because, at the same time, there is an oversupply of aspiring teachers (Chen, 2009; World Bank, 2008).

We use administrative individual-level data from the MoEC on admission criteria and exit examination scores throughout the one-year PPG programme. The data allowed us to analyse the universe of teacher candidates in the 2018 PPG programme for primary school teachers, which included 1,291 teacher candidates. We specifically question whether we can predict a teacher's performance using information available when they were still students. The assumption is that teachers with better admission scores during selection make stronger prospective teachers. We estimated the correlation between the admission test scores and the exit examination scores using an OLS regression conditional on university fixed effects and teacher background. We collected student learning data from 1,530 randomly sampled students taught by 114 teacher candidates approximately two years after the candidates graduated from PPG. Identifying ex-ante criteria that predict teacher subsequent performance would give policymakers a stronger understanding of what programme improvement related to selection criteria may be needed.

We find that the selection criteria significantly predict performance at the end of their PPG programme. The online admission test score strongly predicts a candidate's performance on their knowledge exam. When controlling for teacher characteristics, a one standard deviation higher online admission test score is associated with a 0.30 standard deviation higher knowledge exam score. Performance in the interview during the PPG admission process predicts the candidate's teaching practice exam. Teacher candidates with a one standard deviation higher interview score perform 0.07 standard deviation better on the teaching practice exam. For teacher candidates with one standard deviation higher undergraduate GPA, their knowledge examination performance is 0.15–0.17 standard deviation higher on average, and their teaching practice exam score is 0.06–0.07 standard deviation higher on average.

However, we find no evidence for a correlation between the selection criteria and student learning outcomes. The online admission test does not predict student learning outcomes, either as a composite score or measured separately. Only the undergraduate GPA positively correlates with the student learning outcomes, particularly in numeracy, but the coefficient is small and only significant at the 10 percent level. One standard deviation higher in the undergraduates GPA, on average, is associated with a 0.02 standard deviation increase in student learning outcomes in numeracy. Note that these results may not generalise to other pre-service teachers as our sampled teachers graduated from prominent universities and had substantially higher undergraduate GPAs than the population of candidates in the PPG programme of primary school teacher.

We contribute to the paucity of literature on teacher selection in a developing country context. Few studies on the correlation between teacher selection criteria and learning outcomes in developing countries found mixed results. Bau and Das (2020) found that higher content knowledge was associated with higher teacher value-added in Pakistan, while observed teacher characteristics explained no more

[2] Data from the MoEC's Teachers and Education Personnel (*Guru dan Tenaga Kependidikan*/GTK) Pension Projections, Center for Data and Statistics on Education and Culture (*Pusat Data Dan Statistik Pendidikan Kebudayaan*/PDSPK).

than 5 percent of the variation in teacher value-added. A study by Araujo et al. (2020) used data from Ecuador in an experiment that randomly assigned teachers to teach kindergarten students. The experiment compared learning outcomes of students taught by tenured and contract teachers. The study found that being taught by a tenured teacher who passed a screening test increases students' language scores by 0.1 standard deviation and math scores by 0.08 standard deviation. Another study by Cruz-Aguayo and Ibarrarán (2017), which also used data from Ecuador, found that, at the primary level, there is no correlation between teacher performance during a screening test and their students' learning outcomes.

In the context of the United States, studies on teacher selection point to several selection criteria that predict teacher performance. They found significant correlations for screening scores, such as written assessments, interviews, and sample lessons (Jacob et al., 2018; Hill et al., 2012). Some studies also found a significant correlation for grades during undergraduate study (Jacob et al., 2018; Staiger and Rockoff, 2010), although Kane et al. (2008) found that teachers' undergraduate GPAs were unrelated to student achievement, and Harris and Sass (2011) did not find any evidence that standardised tests for teacher college entrance were related to student achievement. When screening scores were combined into a composite score, as in Rockoff et al. (2011) and Goldhaber et al. (2017), the composite score had more predictive power.

The literature on selection into teacher education programmes mainly focused on teacher candidates' performance at the end of the programme instead of their performance in terms of student learning outcomes after graduation. Admission criteria used to predict candidates' success at the end of the programme includes undergraduate GPA, written profiles, teaching simulation, or standardised tests (Casey and Childs, 2011; Caskey et al., 2001). Both studies found that undergraduate GPA predicted candidates' success at the end of the programme. Another study by Malvern (1991) reported a significant correlation between candidates' interview scores for selection into a postgraduate teacher education programme and their teaching practice exams at the end of the programme.

Our findings suggest that the PPG selection criteria can predict performance of teacher candidates at the end of their education programme, while they cannot predict student learning in a meaningful way. As some other studies did find significant correlations between screening tests and student learning, it is possible that the tests currently used for selection into the PPG programme simply do not measure the right teacher proficiency. For instance, teacher subject knowledge could potentially predict teacher performance better. Still, the limited predictive power of screening tests for student learning in our study suggests that a better way of selecting teachers could be to observe them during a probation (Staiger and Rockoff, 2010) or curation period (Hwa and Pritchett, 2021). In summary, identifying criteria that can predict teacher effectiveness is still an ongoing endeavour.

We organise the rest of this paper as follows. In the next section, we provide background on the PPG programme in Indonesia. In subsequent sections, we describe the data and the method and present our results. The final section contains our conclusion.

2. The PPG Programme

In Indonesia, high school graduates with teaching aspirations can take a four-year degree at a teacher education institution (*Lembaga Pendidikan dan Tenaga Kependidikan/*LPTK).[3] Hereafter, we refer to LPTK as university. After graduating from university, aspiring teachers can apply for teaching jobs. The 2005 Teachers and Lecturers Law stipulates that all teachers must have at least a bachelor's degree and be certified. Certified teachers will receive an additional professional allowance equivalent to their base salary (Chang et al., 2014).[4] It means that certified teachers earn twice as much as non-certified teachers.

Since 2017, teachers who want to obtain certification must attend either of two PPG programmes.[5] Teachers with at least five years of teaching experience can participate in the so called in-service PPG programme for six months (see Figure 1). Meanwhile, teacher college graduates with less than five years of teaching experience can get their certification by participating in the pre-service PPG, a one-year professional education programme. Figure 1 shows that teacher college graduates with different lengths of teaching experience can apply or remain in teaching jobs even if they do not have a teaching certificate.

Although the Teacher Law stipulates that all teachers should be certified, enrolment into these programmes is not mandatory. To encourage enrolment, the Pre-Service PPG programme carries two benefits for aspiring teachers: teacher candidates will receive the teaching certificate once they graduate, so they will receive a higher salary, and they will have an advantage when taking the civil service exam. Teacher candidates are eager to become civil servant teachers because of the financial stability and social respect that comes with the status (Huang et al., 2020). To become a civil servant teacher, teacher candidates need to pass the civil servant exam and they have to apply for a teaching position through the civil service recruitment. PPG graduates will automatically acquire a full score for the field competency selection (*Seleksi Kompetensi Bidang*/SKB) of the civil service exam.

Nevertheless, the status does not warrant the tenured position.[6] PPG graduates are not guaranteed a job. They can apply to either public or private schools. For the hiring at school, PPG graduates benefit from having a teaching certificate but still have to compete with other teacher candidates who do not have teaching certificate.

[3] Teacher education institution or LPTK was known as *Institut Keguruan dan Ilmu Kependidikan* (IKIP). In 1999, IKIP became a university, which means it can open faculties unrelated to teacher education.

[4] In many cases, the certification allowance is a double take-home pay for teachers (Chang et al., 2014). For teachers in private schools, prior to 2020, their take-home pay is set by the school. Based on Government Regulation 6/2020, certification allowance of private school teachers equates to the amount received by civil servant teachers, which is calculated based on a teacher's formal education and years of service.

[5] In the global literature, teacher professional education is generally intended for in-service teachers. However, since 2017, Indonesia differentiates teacher professional education for in-service teachers, called the *PPG Dalam Jabatan*, and for teacher candidates, the *PPG Prajabata*n or Pre-Service PPG.

[6] All teachers passing the civil service recruitment process must undergo two-year probationary period; it is when all recruited civil servant candidates, regardless of their profession, wait for the official announcement of their permanent appointment as a civil servant. This probationary is unrelated to the civil servant candidates' teaching activities in schools; it is primarily about civil servant teachers' responsibility as a government employee rather than a teacher.

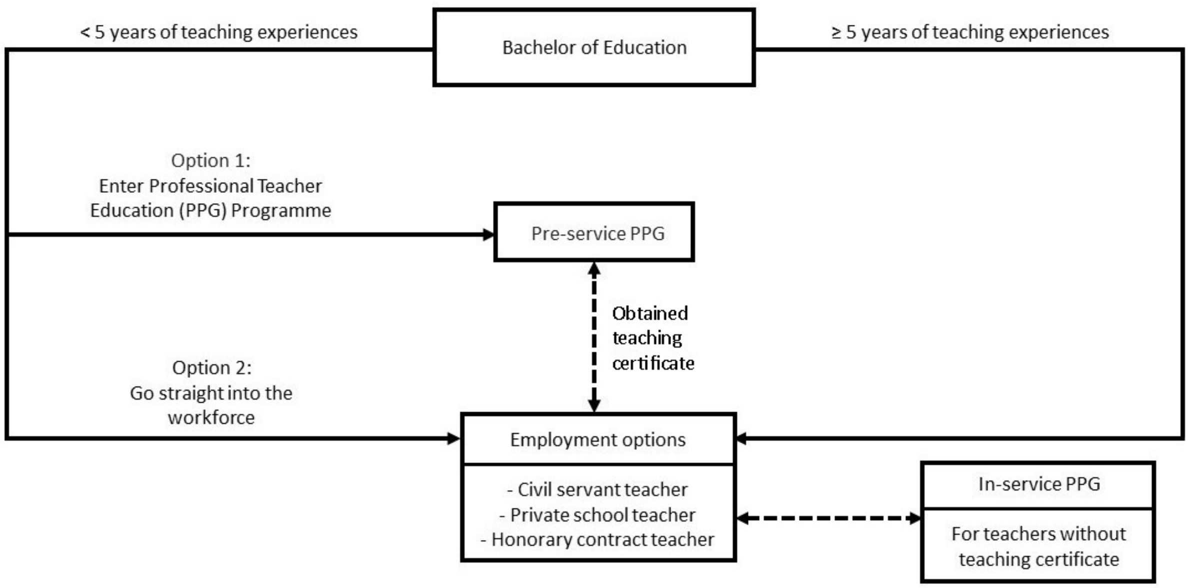

Figure 1. Pathways to Teacher Professional Education Programme and Teaching

The selection procedure of the Pre-Service PPG programme is as follows. To be eligible for programme, an applicant must hold an undergraduate degree with a minimum GPA of 3.0 out of 4.0 from an accredited university. The applicant also must be younger than 31 years old, single, and committed to remain unmarried until completion of the programme. Furthermore, an applicant is allowed to apply if she/he has less than five years of teaching experience before attending the PPG programme.

After passing these administrative requirements, the applicant must go through an online computer-based test. The online admission test is a centralised and standardised exam created by the MoEC. The test measures academic aptitude, English skills, and pedagogical knowledge. Each test score ranges from 0 to 100. The overall online admission test score is a weighted average of the three scores. The tests' weights are as follows: academic aptitude test has a weight of 50%, English skills of 20%, and pedagogical knowledge of 30%. The PPG programme only invites applicants to be interviewed when their score is above 50 on the overall online admission test score.

The interview phase aims to evaluate four aspects of an applicant: their knowledge of the teacher regulation and teacher competencies; their motivation to become a teacher; their personality and talent; and their attitude and appearance (Kementerian Riset, Teknologi, dan Pendidikan Tinggi, 2017b). The interview instrument consists of fifteen indicators to gauge the four aspects. The minimum and maximum absolute scores for the interview are 15 and 60, respectively. The absolute score is then transformed into a percentage that ranges from 25 to 100. During the interview phase, four to five applicants are assessed by two evaluators at the same time. Evaluators selected by the MoEC may ask applicants to conduct microteaching within five to ten minutes or build a discussion among applicants about education in Indonesia.

The applicant then receives a final score, which is the sum of the weighted online admission test and interview scores. The online admission test score is weighted at 70%, while the interview test score is weighted at 30%. Applicants with a final score above 60 are accepted into the Pre-Service PPG programme. They must re-register to confirm their willingness to enrol, after which the MoEC assigns them to selected universities that host the PPG programme. The list of universities that can host the PPG programme is determined by the MoEC.[7] Successful applicants may be assigned to the university where they took their undergraduate education if seats are available. If the university's quota is full, applicants will be transferred to another university that hosts the PPG programme. Applicants cannot choose the university based on their preference.

During the first semester of the PPG programme for primary school teachers, teacher candidates follow a series of in-class workshops, focusing on developing lesson plans for teaching Grades 1 to 6 students, including peer teaching sessions. In the second semester, teacher candidates undergo a period of internship called *Praktik Pengalaman Lapangan* (PPL), where they are assigned to practice teaching at the university's partner primary schools. The internship prepares teacher candidates for the actual learning environment. This arrangement gives them the opportunity to practice the knowledge they gained from the in-class activities, which they did not get during their teaching college.

Upon completing PPG, teacher candidates take a centralised exit exam administered by the MoEC to obtain their teaching certificate. The exam includes knowledge and teaching practice exams. The knowledge exam, *Uji Pengetahuan*, is a written test to assess a teacher candidate's knowledge of pedagogical competence, professional competence, personality, and social competence. The test also includes numeracy, Indonesian language, natural and social science, civic education, methods in developing assessment instruments, the knowledge of differentiated instruction, and school environment.

The teaching practice exam, *Uji Kinerja*, is a competence test to assess a teacher candidate's ability to plan, implement, and evaluate learning. The candidate must provide a lesson plan and prepare learning media used in the classroom three days before the teaching practice exam. Two evaluators in the performance test evaluate the lesson plan, learning media, and the candidate's performance in the classroom. The evaluators are a lecturer and a teacher from the partner school, known as *guru pamong* or mentor, but to avoid favouritism, neither of them is the candidate's supervisor or mentor. The performance test instruments are developed centrally by the MoEC, and the result of the test is uploaded to MoEC's portal. The MoEC provides training for examiners to ensure they have the same perception of the assessment indicators.

We study the teacher candidate cohort that enrolled in February 2018. The passing rate of the exit exam on the first attempt was 94.27% of 1,291 candidates in this cohort. Teacher candidates who failed the exam had another chance to repeat the exam.

[7] MoEC only selects universities with a minimum accreditation of B. Universities in Indonesia are accredited as A (highest level), B, C, or unaccredited.

3. Sample and Data

In this paper, we are interested in the selection mechanism of the Pre-Service PPG programme. We obtained administrative datasets containing the admission test results of applicants who applied for the Pre-Service PPG programme from February to December 2018. We focus on those applicants majoring in primary school teacher education. They enrolled in twenty-nine universities that hosted the PPG programme.

We carried out primary data collection in October 2018 (see Figure 2) on teacher candidates in seven sampled universities regarding their background and prior teaching experience. We purposively selected seven universities in Java Island because once the teacher candidates graduate and are assigned to primary schools, we expected their teaching locations to be in Java. It eases the arrangement of the primary data collection. The seven universities also had larger numbers of PPG students relative to other universities in Java, each about 40 on average. We also did a phone survey in October 2019 to update the information on the employment status and teaching location of our targeted sample.

In January 2020, we attempted to collect data of the targeted sample when they were classroom teachers in the 2019/2020 school year. A classroom teacher or a class teacher at primary schools in Indonesia is responsible for almost everything concerning their class, from taking attendance records of students, teaching all subjects except religion education and physical education to fostering student character. Cases of COVID-19 surged in March 2020 and lockdown measures took place so we could not complete our data collection. We carried out another phone survey in December 2020 to recollect information on the PPG graduates' whereabouts. Based on that second phone survey, we carried out field data collection from February to May 2021. At the time of our visit in 2021, the teacher candidates had graduated from the PPG programme for two years and had been assigned as classroom teachers for six months. Evaluating their performance within two years of graduation is important, as the first two years of novice teachers are when the potential benefits are greatest and retention of teachers is high (Vagi et al., 2017).

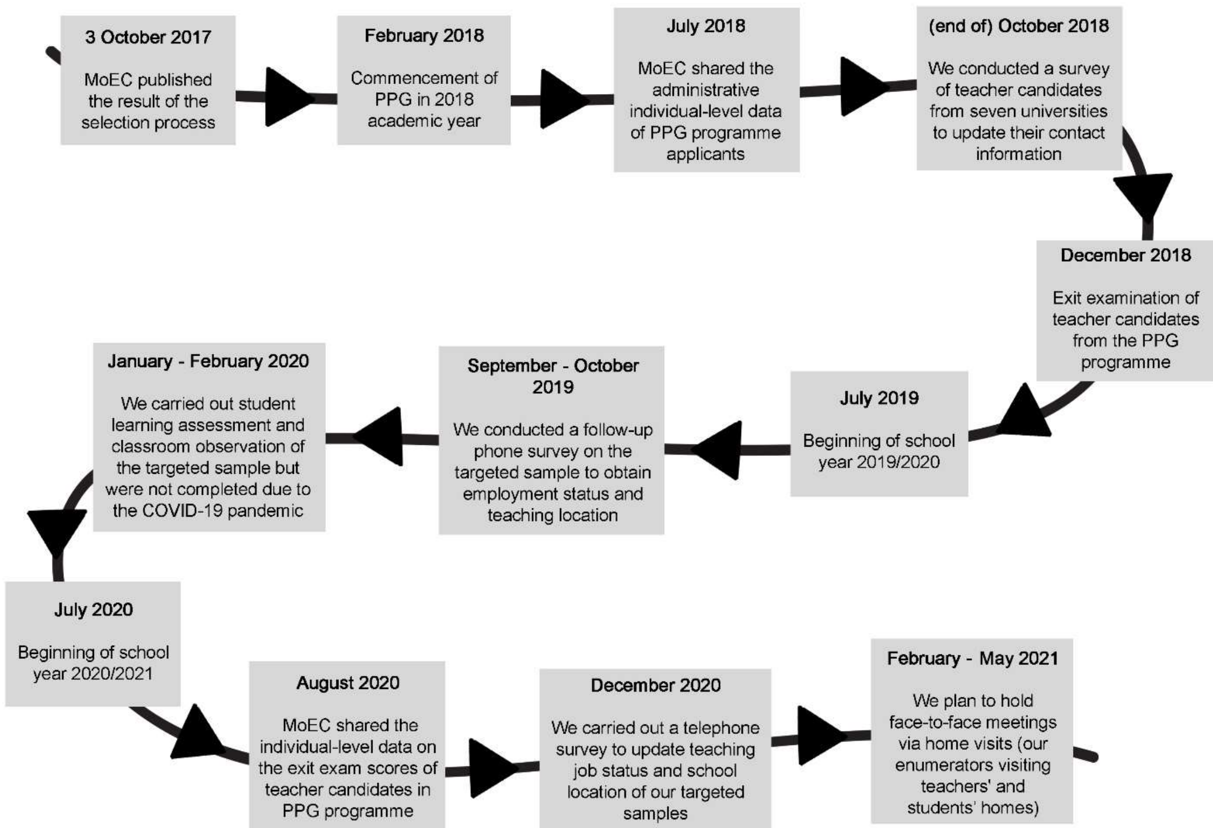

Figure 2. Timeline of the PPG Programme and the Study

We combine the administrative dataset with our primary data collection, as shown in Table 1. The administrative dataset includes data on the accepted and non-accepted applicants. The administrative dataset of the accepted applicants consists of screening measures and exit examination scores. The total number of applicants accepted into the twenty-nine universities that administered PPG programme for primary school teachers were 1,291 teacher candidates. The data covers the total population of applicants.

The six universities on Java sampled for primary data collection had 315 teacher candidates. This dataset is a non-random sample. It consists of all teacher candidates from these seven universities who were visited at the end of their programme and asked to complete our questionnaire. The questionnaire includes data on parents' education and teaching experience prior to PPG. We aimed to visit the 315 teacher candidates once they graduate. However, we were only able to collect data on 110 teachers with the first phone survey. We obtained students tests and classroom observation data from 32 of 110 teachers before the lockdown measures of COVID-19 took place in March 2020, which was followed by school closures.

We carried out another phone survey in December 2020 and included teacher candidates from another university that held pre-service PPG programme.[8] We were able to contact 121 teachers from seven universities. The field data collection was carried out in 2021 and we were able to visit 114 of 121 teachers. We administered student learning assessment on 1,620 students taught by 114 teachers. We interviewed the parents of the students but did not collect data on classroom observation because of the school closures during the COVID-19 pandemic.

Table 1. Data Used in the Analysis

Datasets	Brief Description of Data	Number of Observations	Number of Universities
Admission and exit examination scores data	Data includes names of the accepted applicants, the location of their PPG university, undergraduate GPA, online admission test scores (academic aptitude test, English test, pedagogical knowledge test), interview score, knowledge exam score, performance test score, and passing status.	1,291 teacher candidates	29
Survey data on teacher candidates in six universities	Data includes teacher candidates' characteristics, such as their motivation to join the PPG programme, their teaching experience before joining the PPG programme, their parents' education attainment, and whether they have close relatives who work as teachers or principals.	315 teacher candidates	6
Teacher's classroom observation visited in the 2019/2020 academic year	The classroom observation data includes the frequency of teaching practices in one-time lessons.	32 classroom teachers graduated from PPG	5**
Learning outcomes of students taught by 114 classroom teachers visited in the 2020/2021 academic year	At the student level, the data includes student learning scores in numeracy and literacy tests taken at one time. There is also information on gender, age, parents' education, housing quality index and whether or not the students take private lessons from interview with the parents.	1,620 students of 114 classroom teachers graduated from PPG	7

Notes:
**80% of the 32 sample teachers are from the State University of Jakarta and the Indonesia University of Education Bandung.*

Table 2 below shows descriptive statistics of our dataset. The admission data that we use in the analysis are the GPA in the undergraduate degree, the online admission test scores, and the interview score. The PPG selection requires applicants to hold a minimum GPA of 3 (out of 4) so the variation in our data is limited. Teacher candidates' performance at the end of the PPG programme is measured using their exit examination scores that consists of knowledge and teaching practice exams. We present standard deviation and confidence intervals of the data that we use in the regression analysis to show whether

[8] The additional university, the Indonesia University of Education in Bandung, West Java, was suggested by the MoEC.

variables in the dataset are substantively different from each other between samples. The distribution of data in column (1) is presented in Appendix A.

Table 2. Descriptive Statistics of the Dataset

Variables	(1) Teacher candidates of 29 universities			(2) Teacher candidates of 6 universities			(3) Teachers with student learning outcome data		
	Mean	sd	95% Confidence Interval	Mean	sd	95% Confidence Interval	Mean	sd	95% Confidence Interval
Number of observations	1,291			315			114		
Female	0.74	0.44	(0.72, 0.77)	0.73	0.44	(0.68, 0.78)	0.79	0.41	(0.72, 0.87)
Age	24.69	1.32	(24.62, 24.76)	24.77	1.31	(24.62, 24.91)	24.49	1.21	(24.27, 24.71)
Admission criteria									
Undergraduate GPA	3.56	0.17	(3.55, 3.57)	3.58	0.17	(3.57, 3.60)	3.61	0.17	(3.58, 3.64)
Overall Online admission test Score	56.94	4.89	(56.67, 57.21)	56.34	4.16	(55.88, 56.80)	56.97	4.22	(56.20, 57.75)
- Score on Academic Aptitude Test	71.92	7.18	(71.53, 72.32)	71.41	6.57	(70.68, 72.14)	71.59	6.84	(70.33, 72.85)
- Score on English Test	39.91	8.86	(39.43, 40.39)	38.09	8.24	(37.17, 39.00)	40.26	8.50	(38.70, 41.82)
- Score on Pedagogical Knowledge Test	43.42	6.61	(42.96, 43.68)	43.38	6.86	(42.62, 44.14)	43.76	7.28	(42.42, 45.09)
Interview Score	88.46	7.21	(88.06, 88.85)	88.85	7.04	(88.07, 89.63)	90.99	5.67	(89.95, 92.03)
Exit exam scores									
Knowledge Exam Score	82.37	4.18	(82.15, 82.60)	82.24	3.95	(81.80, 82.67)	82.24	4.11	(81.48, 83.00)
Teaching Practice Exam Score	87.49	4.12	(87.28, 87.71)	86.43	4.36	(86.11,87.08)	88.42	4.84	(87.53, 89.31)

We hypothesise that the online admission test and the interview can measure different teacher qualities. While the online admission test measures cognitive skills, the interview can capture non-cognitive skills. In Appendix B, we show that the correlation between the online admission test score and the interview score is small. Similarly, we find that the correlation between the knowledge and the teaching practice exams is also small.

We collected student test data from 114 classroom teachers who taught Grades 1 to 6 (see Table 3). Since we carried out the data collection during the COVID-19 pandemic, there was no face-to-face teaching. The school closure hindered us to conduct tests of all students taught by the sample teachers. We randomly sampled 15 of the total students taught by the 114 teachers. We administered the student test by visiting each student's house with a strict health protocol in the data collection. In total we tested 1,620 students. Since the test was at student's house, not every parent gave access to the data collection, even when we had made replacements. Similarly, we could only interview parents from 1,530 students to complete information on students' background characteristics. Therefore, the total sample of students to be analysed may not achieved the expected target fifteen students for each teacher.

We collected data on student learning in numeracy and reading using CERMAT (Comprehensive Reading and Numeracy Assessment Tools) developed by Rarasati et al. (2020). The assessments are in accordance with Indonesia's 2006 and 2013 national curricula. These assessment tools are designed to cover a wide range of students' abilities and have a sensitivity to identify improvements in students' literacy and numeracy (see Appendix C). The test for Grades 1 to 3 were carried out orally, via face-to-face meeting with the enumerator, while Grades 4 to 6 were paper based test. We generated the latent ability score of the students using Item Response Theory (IRT).[9] We then standardised the latent ability score based on the mean and standard deviation of students' scores in Grade 1. Table 3 shows the mean and standard deviation of student scores.

Table 3. Descriptive Statistics of Student Scores

Grade	Number of Teachers	Number of Sstudents	Score in Numeracy		Score in Literacy	
			Mean	Std. Dev.	Mean	Std. Dev.
1	11	154	0	1	0	1
2	10	136	0.14	0.80	0.21	0.88
3	17	235	-0.07	0.83	-0.19	0.89
4	22	310	0.02	0.89	0.10	0.98
5	31	439	0.11	1.05	-0.13	0.95
6	23	346	0.25	1.00	0.21	1.00
All grades	114	1620				

Table 3 shows that the test scores do not increase with each grade even though the mean difficulties level of each grade increased proportionately (see Appendix D). This phenomenon can be explained partly by the students' characteristics. In our sample, based on proxy of socio-economic background, the average of student's housing quality index of Grade 2 students is higher, which means economically advantaged compared to the other grade (see Appendix D). While the average housing quality index of Grade 3 students are the lowest, which is likely to be economically disadvantaged compared to the other grade.

To understand the teaching practices carried out by teacher candidates, we used the Classroom Observation Tool for Assessing the Dimensions of Teaching Practices (CERDAS) developed by Yusrina and Bima (2020) to observe the thirty-two teachers while teaching in a classroom. CERDAS is designed to capture teaching practices in the Indonesian context as the instrument uses selected indicators from two national teacher evaluations. CERDAS also includes several indicators in selected international observation instruments, such as structure and type of teacher-student interaction in the classroom used in Trends in International Mathematics Study (TIMSS) Video Study (Ragatz, 2015). Since CERDAS measures the frequencies of teaching practices, it is considered as a low inference instrument.

The classroom observation data allows for analysis of what takes place in the classroom of the thirty-two teachers from PPG graduates. We combined all classroom observation data of the thirty-two teachers to identify common patterns of the lesson features regardless of the subject being taught

[9] Student score generated from IRT method is a measure of latent ability of the students. Students with the same total raw score may have different score because they have different response patterns on the items where each student has a higher probability of answering correctly.

(Indonesian or math). The length of each class varied in the range of 35 to 100 minutes. The observers were to record activities that occurred in the classroom every five minutes. Depending on the length of the lesson, each teacher would have different total observations in the range of 7 to 20 data points. To combine the data from each class, the actual time was not used but rather the relative time from 0 to 100% of class time. For example, for a class of 40 minutes, equal to 8 data points, if teacher's presentation in front of the class occurred for 10 minutes (2 data points) then 25% of the time was used for whole-class interaction. The classes were analysed based on percent of time that passed rather than absolute time.

We analysed six teaching practices from the data we collected using CERDAS (see Table 4). We generate the proportion of each teaching practice by dividing the number of observations intervals marked as one over the total observation intervals of teaching in one lesson. The first three aspects reflect the structure of the teacher-student interaction in the classroom, also shown in Indonesia TIMSS Video Study (Ragatz, 2015). While the latter three aspects are known to be the expected teaching practice in the national teacher evaluation. Table 4 shows that the proportion of instructional time is largely carried out in the whole-class setting. The mean of whole-class setting is 0.350, we say that in average, the whole-class setting constitutes 35% of teachers' instructional time or total observation intervals. Even though evidence of group work setting is said to be the basis of student-centred learning (Ragatz, 2015), the proportion of group work setting is relatively low compared to the other classroom setting.

Table 4. Aspects of Teaching Practices Observed Using CERDAS

Aspects of Teaching Practices	Definition	Mean	Standard Deviation
a. Whole-class setting	If the teacher gives a lecture and presents practice problems in front of the class during observation intervals.	0.35	0.24
b. Group work setting	If the teacher gives group work assignments during observation intervals.	0.14	0.25
c. Individual work setting	If the teacher gives individual assignments during observation intervals.	0.27	0.24
d. Teacher uses learning media	If the teacher uses learning media complementary to the lesson textbook during observation intervals.	0.13	0.23
e. Teacher is making connections between lessons	If the teacher connects today's content with other content taught in the past and/or connects today's content with situations or daily activities commonly experienced by students during observation intervals.	0.03	0.06
f. Teacher asks open-ended questions	If the teacher asks open-ended questions during observation intervals.	0.15	0.19

Notes: Aspects a+b+c should add up to 1 of the total observation intervals. The gap between 1 and (a+b+c) means a proportion of instructional time used by the teacher to transition between class settings.

4. Analytical Methods

We conduct our analysis in three steps, such that we use all information available to us. We standardised the admission criteria and exit exams within this group of 1,291 teacher candidates and use these standardised scores in our estimation. First, we estimate the correlation between the admission criteria (undergraduate GPA, the online admission test scores, and the interview score) and the exit examination scores (the knowledge exam and the teaching practice exam). We begin with the following regression for the sample of 1,291 teacher candidates:

$$Y_i = \alpha_0 + \beta \cdot \boldsymbol{S}_i + \gamma \cdot \boldsymbol{X}_i + \varepsilon_i \quad (1)$$

Where Y is the exit examination scores of teacher candidates *i*, **S** is a vector of undergraduate GPA, online admission test scores and interview score, **X** is a vector of covariates, and ε is an error term. We can estimate the model using the three screening assessments (academic aptitude test score, English test score, and pedagogical knowledge test score) or using its composite score, the overall online admission test score. In the regression for the sample of 1,291 teacher candidates, we only use university fixed effects, i.e., the teacher candidate's undergraduate university, as our control. The administrative dataset that we use in this estimation does not provide control variables.

Second, using the regression form in (1), we estimate the correlation between the admission criteria and the exit examination scores with additional covariates from the survey on the sample of 315 teacher candidates. We add controls for gender, age, motivation, teaching experience, parents' education and for having a family member that is a teacher to the model in equation 1. By including these covariates to the model, we check if the admission scores predict the exit examination scores.

We are interested in whether someone can select the best performing teachers from the pool of applicants, but we only have exit examination scores for the accepted teacher candidates that also enrolled. This could cause selection bias in our estimates. For example, if teacher candidates are accepted based on unobserved characteristics that are correlated with their performance, or if accepted applicants that choose not to enrol have unobserved characteristics that are correlated with their performance. As there is no accepted applicant enrolled with a score below the cut-off, we cannot correct the selection using Heckman model as done in Jacob et al. (2018) and Goldhaber et al. (2017). They use the predicted probability of being in the sample as a function of teacher characteristics. In our case, there is a sharp cut-off value for the screening score, so we can perfectly predict whether or not we observe the exit examination score. However, Jacob et al. (2018) and Goldhaber et al. (2017) find that the correction made little difference.

Third, we estimate the correlation between the screening assessment scores and student learning outcomes, which we collected from the students of 114 teacher candidates who found a teaching position and whom we were able to reach and interview. There are 1,530 observations that we use for the regression below:

$$Z_{si} = \alpha_0 + \beta_1 \cdot \boldsymbol{S}_i + \gamma_1 \cdot \boldsymbol{X}_i + \gamma_2 \cdot \boldsymbol{W}_{si} + \gamma_3 \cdot \bar{Z}_i + \varepsilon_i \quad (2)$$

Where Z is the learning outcome in numeracy and literacy of student *s* of teacher *i*, S is either the admission criteria of teacher *i*, X are teacher-level controls, W are student level controls and $\bar{Z}$ is the classroom average learning score of the students. The observable student characteristics include their grade, gender, parents' education, whether the student is taking private lesson or not, and their housing condition.[10] The classroom average score is the average score of the sampled students in the same classroom. Since the student data consist of one classroom in one school, the average score of the classroom potentially controls for the classroom/peers and school characteristics. Altonji and Mansfield (2018) used the classroom average data of various student's characteristics to correct for sorting of students across teachers. The standard errors in all regressions are clustered at the teacher level.

The model investigates the extent to which admission criteria account for differences in student learning outcomes. There are some limitations which should be kept in mind when interpreting the results. First, our sample only includes the accepted applicants into the PPG programme, therefore, the estimates are applicable only to accepted applicants. Second, using the student learning outcomes at a single point in time may complicate the effort in estimating the effect of our specific interest. The observed student learning assessment score may include error due to different prior ability of students, in which we do not have the available data. However, we can control for some student differences using the available information including students' gender, age, and whether or not they are taking private lessons, housing condition. Lastly, the non-random assignment of teachers to school may attribute to differences in student learning outcomes. Goldhaber (2019) reviewed the literature and showed that teachers prefer to teach in more advantageous schools or in a familiar environment. We include a school quality index in the estimation to address potential bias due to non-random sorting of teachers to schools.

We generated school quality index from the 2017 Ministry of Education and Culture's Education Data Centre (*Data Pokok Pendidikan-Kebudayaan*/DAPODIK 2017). The DAPODIK includes information at the school level regarding school's infrastructure, school's staffing, school's accreditation, and number of students. We use Principal Component Analysis (PCA) to reduce dimension of the selected variables: teacher-student ratio, number of teachers with bachelor's degree, characteristics of classroom with good quality, the provision of library, the provision of science and computer laboratories, and school's accreditation status. However, DAPODIK is only available for schools managed by the MoEC.

We complete our analysis on the correlation between PPG selection criteria and teacher performance by making use of classroom observation data from thirty-two PPG graduate teachers. Following the TIMSS video study analysis, we assume that teacher influences the learning process through their teaching practices in the classroom (Ragatz, 2015). Similarly, we assume that teachers with higher levels of admission scores (i.e., the online admission, interview test scores, and undergraduate GPA) could tend to use different practices than those with lower levels. Since we only have classroom observation data of 32 thirty-two teachers, we differentiate them as those in the top on-third and bottom on-third of distribution of each admission criteria. It is important to note that there might be other teacher characteristics or students' background, grade level, and class size that drives teacher's teaching practice in the classroom.

Other than the small number of sample teachers, the limitation of the data used in this analysis is that we only capture teaching practice at one time and based on live observation of one observer/rater. Thus, it may not reflect variation in individual practices of a teacher that occur over the course of a school year.

[10] The data on housing condition includes type of roof, wall, floor, cooking fuel, improved water source, improved sanitation, and availability of electricity. We use Principal Component Analysis (PCA) to reduce dimension of the selected variables and construct housing quality index.

5. Results

5.1. Correlation between Admission Criteria and the PPG Exit Exams

The results of the estimation of equation (1) are shown in Table 4. We use data of all teacher candidates in pre-service PPG programme in primary school in twenty-nine universities. We find that the academic aptitude test score, English test score, and pedagogical knowledge test score are significantly correlated with the knowledge exam result (Column 1). One standard deviation higher overall online admission test score is associated with a 0.3 standard deviation higher knowledge exam score (Column 3) and a one standard deviation higher interview score is associated with a 0.07 standard deviation higher teaching practice exam score (Column 6).

Of the online admission test components, the pedagogical knowledge test score and the academic aptitude test score have the largest correlation with the knowledge exam score. This makes sense because there were no English test items in the knowledge exam. The undergraduate GPA is correlated with both exit exams. Teacher candidates with undergraduate GPA of one standard deviation higher scores on average 0.15 standard deviation higher on the knowledge exam (Column 2) and 0.07 standard deviation higher on the teaching practice exam (Column 5).

Our results suggest that the online screening tests predict the knowledge exam score, while they are not associated with the teaching practice exam score.

Table 4. Regression Results between Admission Criteria and Exit Examination

	(1)	(2)	(3)	(4)	(5)	(6)
	Standardised Score on Knowledge exam			Standardised Score on Teaching practice exam		
Standardised Score on Academic Aptitude Test	0.22^{***}	0.18^{***}		0.01	0.04	
	(0.03)	(0.03)		(0.03)	(0.02)	
Standardised Score on English Test	0.08^{***}	0.06^{**}		0.00	0.01	
	(0.03)	(0.03)		(0.03)	(0.02)	
Standardised Score on Pedagogical Knowledge Test	0.29^{***}	0.24^{***}		-0.03	0.01	
	(0.03)	(0.03)		(0.03)	(0.02)	
Standardised Overall Online admission test Score			0.30^{***}			0.04
			(0.03)			(0.03)
Standardised Interview Score	0.06^{**}	0.02	0.03	0.10^{***}	0.07^{***}	0.07^{***}
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Standardized Undergraduate GPA		0.15***	0.17***		0.07***	0.06***
		(0.03)	(0.03)		(0.02)	(0.02)
Constant	-0.00	0.41**	0.42**	-0.00	0.42***	0.43***
	(0.03)	(0.21)	(0.21)	(0.03)	(0.15)	(0.15)
LPTKs dummy	No	Yes	Yes	No	Yes	Yes
Observations	1291	1291	1291	1291	1291	1291
R^2	0.16	0.32	0.30	0.01	0.53	0.53

Note: Standard errors in parentheses and robust. Note that GPA of the selected students ranges between 3 and 4 and we standardized within this group of 1291 individuals. * $p < .10$, ** $p < .05$, *** $p < .01$

5.2. Correlation between Admission Criteria and the PPG Exit Exam with Additional Control Variables from Survey

Using the survey sample of 315 teacher candidates[11], we add background control variables to the model in equation (1). We find very little correlation between background characteristics and the online admission test and interview scores. We only find that students with prior teaching experience score lower on the online admission test (Column 1).

The correlations between the admission criteria and the exit exam scores are similar in this sample and adding the background characteristics as controls does not alter the findings. The criterion of online admission test score only predicts candidates' performance in the knowledge exam while the interview score only predicts teaching practice exam. The coefficient for undergraduate GPA is somewhat smaller (Column 3 and 4), and no longer significant for the teaching practice exam (Column 5 and 6).

We find that candidates with prior teaching experience score significantly higher on the teaching practice exam by 0.4 standard deviation, when controlling for other factors (Column 6). We also find that candidates whose fathers have a higher education level score 0.31 standard deviation higher on the teaching practice exam compared to those whose fathers did not complete high school, when controlling for other factors (Column 6). Based on Wang et al. (2020), parents' education correlates significantly with non-cognitive ability of their children in China. Higher mother's education correlates negatively with most of emotional behaviour of the children while father's education correlates positively with the number of friends their children have. Wang et al.'s (2020) finding provides insight that the higher the father's education, the more likely the children interact or be able to communicate with other people. It suggests that candidates with a father who completed higher education are likely to perform better in doing teaching in front of others.

[11] We excluded one observation from the sample in the regression since the two respondents had missing values in the control variables.

Table 5. Regression Results between Teacher Candidates' Characteristics, Admission Criteria, and Exit Examination

	(1) Standardised Online admission test Score	(2) Standardised Interview Score	(3) Standardised Score on Knowledge exam	(4)	(5) Standardised Score on Teaching practice exam	(6)
Standardised Online admission test Score			0.36*** (0.06)	0.39*** (0.06)	0.05 (0.05)	0.07 (0.05)
Standardised Interview Score			0.04 (0.07)	0.02 (0.07)	0.11** (0.05)	0.12** (0.05)
Standardized Undergraduate GPA	0.01 (0.06)	0.11** (0.06)	0.12** (0.05)	0.14** (0.06)	0.01 (0.04)	0.00 (0.05)
Female	-0.10 (0.12)	0.21 (0.13)		0.09 (0.13)		0.06 (0.10)
Age	0.06 (0.04)	-0.04 (0.04)		0.03 (0.04)		-0.04 (0.03)
Joined PPG to improve teaching competencies	0.06 (0.10)	0.04 (0.11)		-0.01 (0.13)		-0.11 (0.10)
Teaching experience prior to PPG	-0.36* (0.20)	0.01 (0.17)		0.35** (0.14)		0.40*** (0.14)
Father completed high school	-0.12 (0.14)	0.13 (0.14)		0.18 (0.14)		0.13 (0.11)
Father completed higher education	0.16 (0.14)	-0.05 (0.15)		-0.12 (0.16)		0.31** (0.12)
Mother completed high school	-0.09 (0.12)	0.02 (0.12)		-0.21 (0.13)		0.02 (0.11)
Mother completed higher education	-0.01 (0.14)	-0.19 (0.15)		-0.20 (0.14)		-0.03 (0.11)
Family member with teaching	-0.03 (0.11)	0.08 (0.12)		-0.03 (0.12)		0.00 (0.10)

	(1)	(2)	(3)	(4)	(5)	(6)
job						
Constant	-1.18	0.37	-0.51***	-1.47	0.51***	1.61*
	(1.04)	(0.96)	(0.14)	(0.98)	(0.12)	(0.85)
LPTKs dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	315	315	314	315	314	315
Adjusted R^2	0.16	0.30	0.17	0.20	0.52	0.53
R^2	0.20	0.33	0.20	0.24	0.53	0.55

Note: Columns 1 and 2 do not correct for the fact that the admission scores are bounded by a minimum threshold. Standard errors in parentheses and robust. Note that GPA of the selected candidates range between 3 and 4 and we standardized within this group of 1291 individuals. * $p < .10$, ** $p < .05$, *** $p < .01$

5.3. Correlation between Admission criteria and Student Learning Assessment in Standardised Math and Literacy Test

In this section, we focus on all students taught by 114 graduates from the PPG programme in seven universities. They had been assigned to or hired at primary schools as classroom teachers in the academic year of 2020/2021 for approximately eight months. Our dataset consists of 1,620 students taught by the 114 teachers, but only 1,530 students had complete information on students' background characteristics. We estimate Equation 2 to find the correlation between the admission criteria of teacher candidates with student scores in numeracy and literacy. We include a school quality index and the classroom average test score to account for the school and classroom effect. We do not have baseline test scores of students to control for student prior knowledge. To the extent of available data, we include student characteristics to account for students' background.

In Table 6, we observe the correlation between teachers' admission criteria and the students' numeracy scores. The results in the Table 6 show an insignificant correlation between teachers' online admission test, interview performance, undergraduate GPA, and students' numeracy skills (Column 1). When we control for teacher and students background characteristics (Column 2), the admission criteria are not significant and variation in student learning outcomes mostly explained by students' characteristics.

In column 3, we then include the classroom average score and the school quality index into the model. We find that in such specification, the undergraduate GPA become significant at 10 percent level. For one standard deviation increased in the undergraduate GPA, the students' outcome in numeracy test increased by 0.02, when controlling for teacher and students background characteristics. The adjusted *R*-squared values that explain the variance in teacher performance of accepted applicants is also relatively larger when we include both, classroom and school effect, in the specification.

The undergraduate GPA positively correlated with students' numeracy outcomes when we use the academic aptitude test, English test, and pedagogical knowledge test in the specification (column 5). However, when we did not include the university where the teachers took their PPG programme in the specification (Column 4 and 6), the undergraduate GPA is not significant. The results suggest that the university origin of the teachers might drive differences in the undergraduate GPA that affecting students learning outcome in numeracy. On that note, teachers with the same GPA points might be different in terms of the quality if they graduate from different university.

Table 6. Regression Results on Admission Criteria and Student Numeracy Skills

	Standardised Score on Numeracy Test					
	(1)	(2)	(3)	(4)	(5)	(6)
Standardised Score on Online Admission Test	-0.07	-0.08	-0.01	-0.01		
	(0.09)	(0.08)	(0.01)	(0.01)		
Standardised Score on Academic Aptitude Test					-0.01	-0.00
					(0.01)	(0.01)
Standardised Score on English Test					0.00	-0.01
					(0.01)	(0.01)
Standardised Score on Pedagogical Knowledge Test					-0.00	0.00
					(0.01)	(0.01)
Standardised Score on Interview	-0.10	-0.11	-0.00	-0.00	-0.00	-0.00
	(0.10)	(0.10)	(0.01)	(0.01)	(0.01)	(0.01)
Standardized Undergraduate GPA	0.02	0.05	0.02*	0.01	0.02*	0.01
	(0.07)	(0.07)	(0.01)	(0.01)	(0.01)	(0.01)
Teacher is female		0.14	-0.02	-0.02	-0.02	-0.02
		(0.16)	(0.02)	(0.02)	(0.02)	(0.02)
Teacher's age		-0.00	-0.01*	-0.01	-0.01	-0.01
		(0.07)	(0.01)	(0.01)	(0.01)	(0.01)
Teacher's years of teaching prior to PPG		0.03	0.01	0.01	0.01	0.01
		(0.07)	(0.01)	(0.01)	(0.01)	(0.01)
Student is female		0.11***	0.09**	0.09***	0.09**	0.09**
		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Student is taking private lesson		0.23**	0.08	0.07	0.08	0.08
		(0.11)	(0.07)	(0.06)	(0.07)	(0.06)
Student’s father attended college		0.24**	0.02	0.02	0.02	0.02
		(0.10)	(0.07)	(0.07)	(0.07)	(0.07)
Student’s mother attended college		0.26**	0.09	0.09	0.09	0.09
		(0.10)	(0.08)	(0.08)	(0.08)	(0.08)
Student's housing quality index		0.07***	0.03**	0.03**	0.03**	0.03**
		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
School's quality index			0.00	-0.00	0.00	-0.00
			(0.00)	(0.00)	(0.00)	(0.00)
Classroom average score in Numeracy			0.97***	0.97***	0.97***	0.97***
			(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.00	-0.40	0.20	0.20	0.19	0.20
	(0.32)	(1.68)	(0.19)	(0.20)	(0.20)	(0.21)
LPTKs Dummy	Yes	Yes	Yes	No	Yes	No
Grade Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1620	1530	1530	1530	1530	1530
R^2	0.05	0.11	0.49	0.49	0.49	0.49
Adjusted R^2	0.04	0.09	0.48	0.48	0.48	0.48

Note: Standard errors in parentheses and clustered at the teacher level. * $p < .10$, ** $p < .05$, *** $p < .01$. Note that GPA of the selected candidates' range between 3 and 4 and we standardized within the group of 1291 individuals.

The correlation between the admission criteria and the students' literacy test scores are shown in Table 7. In column 1, the standardised score of interview shows a negative correlation with students learning outcome in literacy. The significance levels of interview score tended to drop when including teacher and student background characteristics, while the standardised undergraduate GPA correlates positively with students' outcome in literacy (Column 2).

The undergraduate GPA continues to be a significant predictor of student learning outcomes in literacy when we also include school's quality and classroom average score in the estimation (Column 3). However, when we use academic aptitude test, English test, and pedagogical knowledge test in the specification, the undergraduate GPA is no longer significant (Column 5). The results suggest that the use of undergraduate GPA to predict for student learning outcomes in literacy were mixed.

Using the same steps of robustness check as in the estimation of student numeracy test, when we did not account for university of origin, the undergraduate GPA in the estimation of student literacy test is not significant (Column 4 and 6). It might also shows as evidence of different grading standards across universities.

Following Rockoff et al. (2011) and Goldhaber et al. (2017), we generated a composite score of admission criteria using principal component analysis. We found that the composite score, which includes the online admission, interview test scores, and undergraduate GPA), also did not result in meaningful correlations with student learning.

Table 7. Regression Results on Admission Criteria and Student Literacy Skills

	Standardised Score on Literacy Test					
	(1)	(2)	(3)	(4)	(5)	(6)
Standardised Score on Online Admission Test	-0.06	-0.07	0.00	-0.00		
	(0.07)	(0.06)	(0.01)	(0.00)		
Standardised Score on Academic Aptitude Test					-0.00	-0.00
					(0.01)	(0.01)
Standardised Score on English Test					0.00	0.00
					(0.01)	(0.00)
Standardised Score on Pedagogical Knowledge Test					0.01	0.01
					(0.01)	(0.01)
Standardised Score on Interview	-0.14*	-0.10	0.00	0.00	0.00	0.00
	(0.07)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)
Standardized Undergraduate GPA	0.05	0.10*	0.02*	0.01	0.01	0.01
	(0.05)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)
Teacher is female		-0.14	-0.03	-0.02	-0.03	-0.02
		(0.13)	(0.02)	(0.02)	(0.02)	(0.02)
Teacher's age		0.04	-0.00	-0.00	-0.00	-0.00

		(0.06)	(0.00)	(0.00)	(0.00)	(0.00)
Teacher's years of teaching prior to PPG		-0.01 (0.05)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Student is female		0.21*** (0.04)	0.18*** (0.04)	0.18*** (0.04)	0.18*** (0.04)	0.18*** (0.04)
Student is taking private lesson		0.18** (0.09)	0.05 (0.06)	0.05 (0.05)	0.05 (0.06)	0.05 (0.05)
Student’s father attended college		0.20** (0.09)	0.01 (0.07)	0.01 (0.07)	0.01 (0.07)	0.01 (0.07)
Student’s mother attended college		0.26** (0.12)	0.12 (0.10)	0.12 (0.10)	0.12 (0.10)	0.12 (0.10)
Student's housing quality index		0.07*** (0.03)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
School's quality index			-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Classroom average score in Literacy			0.97*** (0.01)	0.97*** (0.01)	0.97*** (0.01)	0.98*** (0.01)
Constant	0.12 (0.27)	-1.20 (1.44)	-0.09 (0.11)	-0.08 (0.11)	-0.11 (0.11)	-0.10 (0.11)
LPTK Dummy	Yes	Yes	Yes	No	Yes	No
Grade Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1620	1528	1528	1528	1528	1528
R^2	0.08	0.13	0.34	0.34	0.34	0.34
Adjusted R^2	0.07	0.11	0.33	0.33	0.33	0.33

	Standardised Score on Literacy Test					
	(1)	(2)	(3)	(4)	(5)	(6)
Standardised Score on Online Admission Test	-0.06 (0.07)	-0.07 (0.06)	0.00 (0.01)	-0.00 (0.00)		
Standardised Score on Academic Aptitude Test					-0.00 (0.01)	-0.00 (0.01)
Standardised Score on English Test					0.00 (0.01)	0.00 (0.00)
Standardised Score on Pedagogical Knowledge Test					0.01 (0.01)	0.01 (0.01)
Standardised Score on Interview	-0.14* (0.07)	-0.10 (0.06)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Standardized Undergraduate GPA	0.05 (0.05)	0.10* (0.05)	0.02* (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Teacher is female		-0.14 (0.13)	-0.03 (0.02)	-0.02 (0.02)	-0.03 (0.02)	-0.02 (0.02)
Teacher's age		0.04 (0.06)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Teacher's years of teaching prior to PPG		-0.01 (0.05)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)

Student is female		0.21***	0.18***	0.18***	0.18***	0.18***
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Student is taking private lesson		0.18**	0.05	0.05	0.05	0.05
		(0.09)	(0.06)	(0.05)	(0.06)	(0.05)
Student's father attended college		0.20**	0.01	0.01	0.01	0.01
		(0.09)	(0.07)	(0.07)	(0.07)	(0.07)
Student's mother attended college		0.26**	0.12	0.12	0.12	0.12
		(0.12)	(0.10)	(0.10)	(0.10)	(0.10)
Student's housing quality index		0.07***	0.03	0.03	0.03	0.03
		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
School's quality index			-0.00	-0.00	-0.00	-0.00
			(0.00)	(0.00)	(0.00)	(0.00)
Classroom average score in Literacy			0.97***	0.97***	0.97***	0.98***
			(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.12	-1.20	-0.09	-0.08	-0.11	-0.10
	(0.27)	(1.44)	(0.11)	(0.11)	(0.11)	(0.11)
LPTK Dummy	Yes	Yes	Yes	No	Yes	No
Grade Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1620	1528	1528	1528	1528	1528
R^2	0.08	0.13	0.34	0.34	0.34	0.34
Adjusted R^2	0.07	0.11	0.33	0.33	0.33	0.33

Note: Standard errors in parentheses and clustered at the teacher level. * $p < .10$, ** $p < .05$, *** $p < .01$. Note that GPA of the selected candidates range between 3 and 4 and we standardized within the group of 1291 individuals.

5.4. Teaching Lesson Patterns from the PPG Graduates Based on the Classroom Observation Data

In the previous section, the correlation of online admission test score, interview score, and the student learning outcomes yields insignificant results. We seek further if the insignificant results may come from the fact that there are no differences in the teaching practice of the teachers across distribution of their selection criteria. The objective of the analysis is to examine teaching practices across the selection criteria. We want to know descriptive differences in the classroom setting or teaching practices between teachers scoring in the top one-third and bottom on-third of the online admission score and interview score. We are interested to know whether teachers with different scores use different practices.

The whole-class setting was by far the most common form of classroom setting across all sample, with group work setting being less common (see Figures 3 to 5). The whole-class setting includes lecturing, explaining students' assignments, as well as providing example problems and the solution. The transition between classroom settings occurs about 24% of the lesson time across all sample. The transition means a switch from whole-class to group work or individual work and vice versa that occurs within five minutes of the observation. The more percentage of lesson time is used for transition means less time are being used for effective teaching instruction.

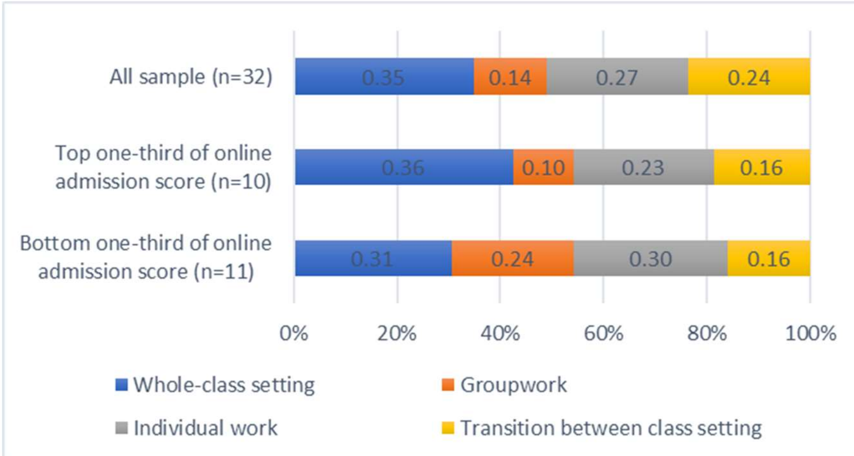

Figure 3. Types of Classroom Setting Based on Teacher's Online Admission Score

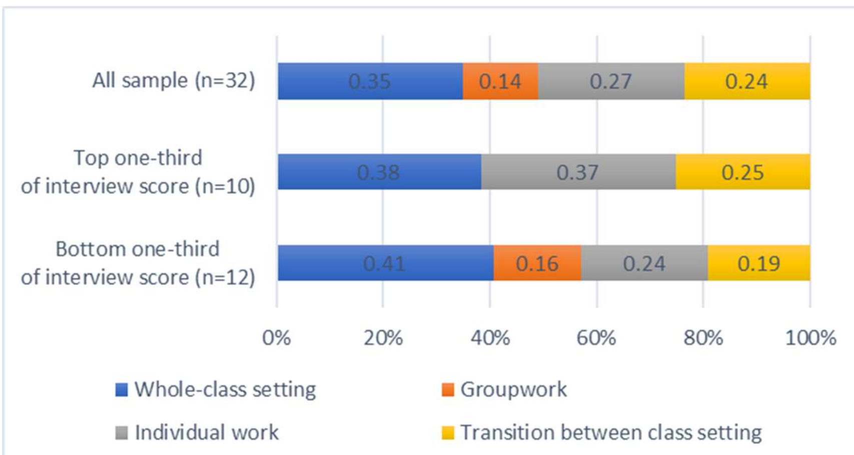

Figure 4. Types of Classroom Setting Based on Teacher's Interview Score

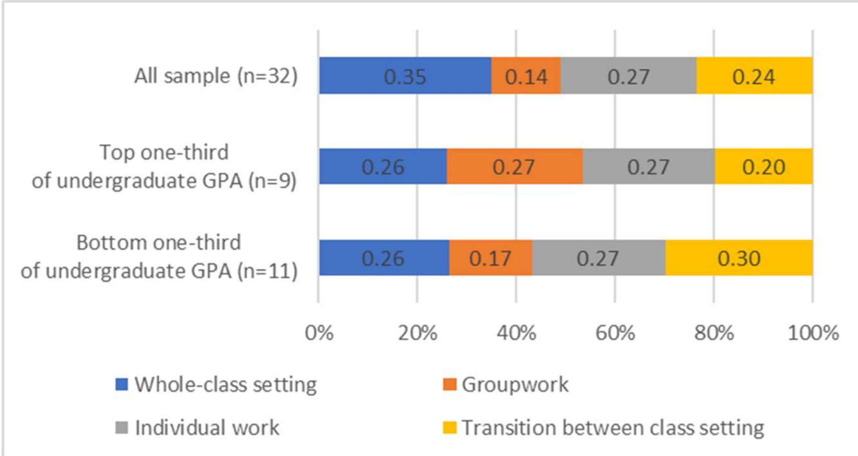

Figure 5. Types of Classroom Setting Based on Teacher's Undergraduate GPA

We find a similar pattern of teacher classroom settings based on the distribution of online admission tests and interview scores. Teachers in the top one-third of distribution in online admission and interview scores tend to use less group work settings than the teachers in the bottom one-third of each score distribution.

However, when we differentiate teacher classroom settings based on their undergraduate GPA, those in the top one-third are likely to use more group work setting than teachers in the bottom one-third. Teachers in the top one-third of undergraduate GPA distribution also seems to spend less in doing transition.

To gauge teacher teaching practice that boosts students' understanding (Bruns and Luque, 2014; Siraj and Taggart, 2014; Lee and Kinzie, 2011), we analyse the use of learning materials and the occurrence of teacher's effort in making a connection between lessons and asking open-ended questions. These teaching practices may occur in any classroom setting since the observation instrument captures several activities during the five-minute observation interval.

Teachers only use learning materials other than compulsory textbooks for 13% of the lesson time in the entire sample (see Figure 6). That means, for a class of 40 minutes, which equals 8 data points, the teacher uses learning materials for about 5 minutes. Teachers in the bottom one-third of online admission scores, interview scores, and undergraduate GPA are likely to spend more time using learning materials than teachers in the bottom one-third of each distribution (see Figures 6, 7, and 8). Unfortunately, our classroom observation data cannot capture the types of learning materials used. However, teacher's use of learning materials, whether technology-based or not, keeps students engaged in learning (Bruns and Luque, 2014).

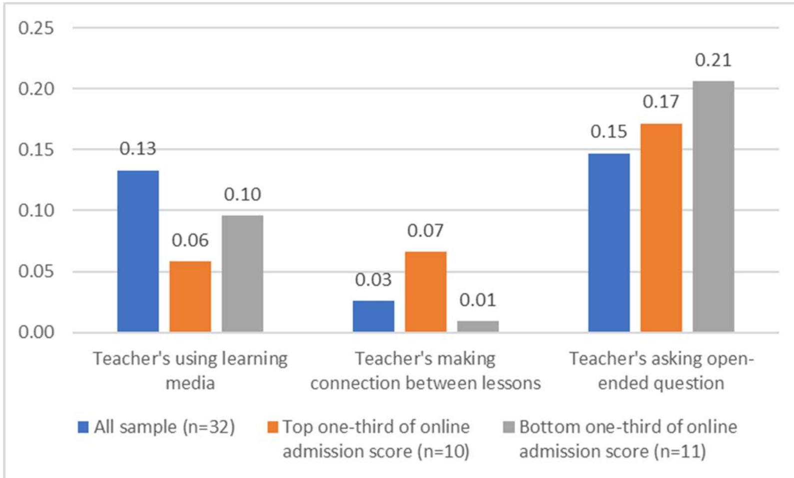

Figure 6. Teaching Practices Based on Teacher's Online Admission Score

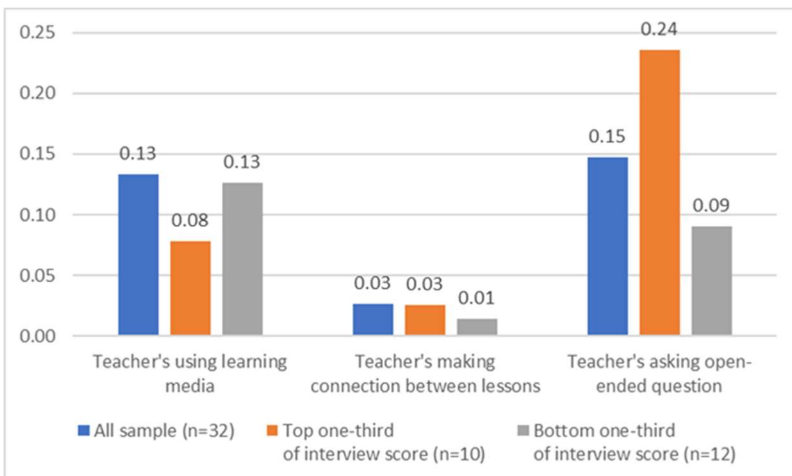

Figure 7. Teaching Practices Based on Teacher's Interview Score

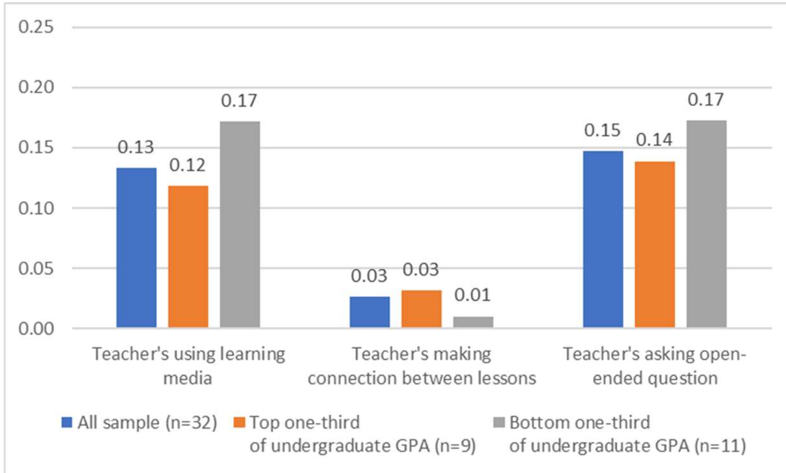

Figure 8. Teaching Practices Based on Teacher's Undergraduate GPA

Teacher efforts in making connection in teaching rarely occurs during lesson across teachers in the distribution of each admission criteria. When in fact, teachers who are connecting their lesson with other subjects, previous contents or daily-life information can strengthen students' understanding of the content (Siraj and Taggart, 2014).

From Figures 6 to 8, the graphs show that the proportion of time teachers spend asking open-ended questions is relatively higher than in the previous two teaching practices. It accounts for 15% of the lesson time across all samples. Teachers in the top one-third of interview scores seem to spend more time asking open-ended questions. Questioning is an attempt by a teacher to check student understanding (Ragatz, 2015). When teachers ask open-ended questions, it elicits students' thinking rather than simply recalling information (Lee and Kinzie, 2011).

We found no clear patterns in teaching practices across teachers in the distribution of online admission test scores, interview scores, and undergraduate GPA. The result shows that teachers in the top one-third of distribution in online admission test scores, interview scores, and undergraduate GPA do not necessarily will exert teaching practices that we assume will gauge students' understanding.

6. Discussion

The study evaluates whether the admission criteria used by the PPG programme are predictive of candidates' exit exam and actual classroom performances. We find that undergraduate GPA, online admission test, and interview predict a teacher candidate's success at the end of the PPG programme. A one standard deviation higher score on the online admission test, which includes the scores of academic aptitude test, English test, and pedagogical knowledge test, is associated with a 0.3 standard deviation higher score on the knowledge exit exam, also when conditioning on teacher background characteristics. A one standard deviation higher score on the interview is associated with a 0.07 standard deviation higher score on the teaching practice exam.

We found that the overall online admission test score indeed predicts a candidate's success in the knowledge exam, whereas the interview only predicts their success in the teaching practice exam. The online admission test correlates positively with knowledge exam because the underlying construct of both are the academic ability of teachers. Our result for the correlation between interview and teaching practice exam is similar to Malvern (1991) who found that a selector's judgment in an interview correlates with a candidate's teaching practice performance abilities. The author argued that the interview typically looks for personality traits representing teaching abilities valued in the practice exam rather than academic abilities. Therefore, it is likely that the interview only predicts the teaching practice exam score and cannot predict the knowledge exam score.

The undergraduate GPA is predictive for both exit examinations, although it is restricted to above 3 (and maximum 4) for all admitted candidates. A candidate with a one standard deviation higher in undergraduate GPA, on average, scores 0.17 one standard deviation higher on the knowledge exam and 0.06 standard deviation higher on the practice exam. Our result follows Caskey et al. (2001) in that the undergraduate GPA shows the lowest correlation with candidate's exit exams compared to other admission criteria. We also found that the positive correlation between undergraduate GPA and the teaching practice exam is smaller than that of the knowledge exam. It also means that the undergraduate GPA is just as good for predicting teaching practice scores as predicting the knowledge exam score.

While we found that all admission criteria are associated with success at the end of the PPG programme, the next question is whether the PPG selection criteria lead to an improvement in the performance of the teacher workforce in terms of student learning outcomes. Our result shows that only the undergraduate GPA correlates positively with student learning outcomes, particularly numeracy. However, the coefficient is small and only significant at the 10 percent level. A one standard deviation higher in undergraduate GPA, on average, scores 0.02 one standard deviation higher on the numeracy test. The significant result of undergraduate GPA should be taken with caution since it cannot predict student learning outcomes in a meaningful way.

In terms of the pedagogical knowledge test, our results contradict evidence from the Indonesian TIMSS video study, which shows that such a test predicts student learning outcomes (Ragatz, 2015). The TIMSS study noted that translating the pedagogical concept into a written test presents difficulties. Due to such difficulties, there is a chance that the pedagogical knowledge test cannot gauge teacher performance when administered in the selection process.

We complement our estimation of the correlation between admission criteria and teacher performance with an analysis of classroom observation data of thirty-two PPG graduate teachers. We found no clear patterns in teaching practices across teachers in the distribution of online admission test scores, interview scores, and undergraduate GPA. This finding is similar to Ragatz (2015), who argued that higher scores teachers did not always use different practices than those at the lower levels. Ragatz indicated that teachers might use the same practices, but some more effectively. Although there are no patterns in teaching practices across the distribution of admission criteria, it is also possible that sample teachers of PPG graduates have yet to perform a well-defined way of teaching at the time of the observation. One should note that whether the classroom practices influence student performance is beyond our analysis.

7. Conclusion and Policy Recommendation

We conclude that PPG admission criteria can be used as ex-ante criteria to predict teacher performance at the end of the programme, but they do not predict performance in the classroom. A selection process based on the current measures is likely to only reduce the total number of applicants to be screened without necessarily identifying individuals with the potential to be effective teachers. At this stage, it might not replace the soon retired teachers with better-performing teachers.

Barber and Mourshed (2007) mentioned that in a high-performing education system, the selection of teachers puts a strong emphasis on a high level of numeracy and literacy, interpersonal and communication skills, and motivation to teach. To the extent that the current PPG admission criteria cannot predict student learning outcomes, the PPG programme admission can potentially refine their admission process by including subject content knowledge test, particularly in numeracy and literacy content. Previous studies argued that the subject content knowledge might serve as a more stable predictor of candidates' success (Caskey et al., 2001) and a better predictor of student learning outcomes (Ragatz, 2015). In Peru and Pakistan, Metzler and Woessmann (2012) and Bau and Das (2020) show that subject-specific teacher achievement correlates with student achievement. Meanwhile, the pedagogical content knowledge test can also be considered to improve the screening process (Hwa and Pritchett, 2021).

Selection of teachers is just one phase in the chain of an education system. Mixed results from the previous studies on the predictive ability of ex-ante criteria on teacher performance (Klassen and Kim, 2019) and our study suggest that using the selection alone to identify effective teachers may not suffice and should be combined with continuous screening over teachers' career (Braga et al., 2020). The strategy could be in the form of a probationary period (Staiger and Rockoff, 2010) or a period of curation (Hwa and Pritchett, 2021) before becoming in-service teachers or professional development activities during the in-service (Braga et al., 2020). But that does not mean that the selection is less important. There is also the benefit to investing in a rigorous selection process, as it reduces the cost of such a strategy (Hwa and Pritchett, 2021). More broadly, this study highlights the need for further research on the teacher selection criteria that best predict teacher performance in actual classroom settings.

References

Ackley, B., Fallon, M. and Brouwer, N. (2007). Intake Assessments for Alternative Teacher Education: Moving from Legitimation towards Predictive Validity. *Assessment & Evaluation in Higher Education*, [online] 32(6), pp.657-665. Available at: <https://www.tandfonline.com/doi/abs/10.1080/02602930601117134>.

Altonji, J. and Mansfield, R. (2018). Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects. *American Economic Review*, [online] 108(10), pp.2902-2946. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20141708>.

Araujo, M., Heineck, G. and Cruz-Aguayo, Y. (2020). *Does Test-Based Teacher Recruitment Work In The Developing World? Experimental Evidence from Ecuador*. IZA DP No. 13830. [online] Bonn: IZA Institute of Labor Economics. Available at: <https://www.iza.org/publications/dp/13830/does-test-based-teacher-recruitment-work-in-the-developing-world-experimental-evidence-from-ecuador>.

Barber, M. and Mourshed, M. (2007). *How the World's Best-Performing School Systems Come Out on Top*. [online] McKinsey & Company. Available at: <https://www.mckinsey.com/industries/education/our-insights/how-the-worlds-best-performing-school-systems-come-out-on-top>.

Bau, N. and Das, J. (2020). Teacher Value Added in a Low-Income Country. *American Economic Journal: Economic Policy*, [online] 12(1), pp.62-96. Available at: <https://www.aeaweb.org/articles?id=10.1257/pol.20170243>.

Braga, M., Checchi, D., Scervini, F. and Garrouste, C. (2020). Selecting or Rewarding Teachers? International Evidence from Primary Schools. *Economics of Education Review*, [online] 76, p.101986. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0272775719301992?via%3Dihub>.

Bruns, B. and Luque, J. (2014). *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean (Advance Edition)*. [ebook] Washington, D.C.: World Bank. Available at: <https://openknowledge.worldbank.org/handle/10986/20488> [Accessed 8 April 2018].

Casey, C. and Childs, R. (2011). Teacher Education Admission Criteria as Measure of Preparedness for Teaching. *Canadian Journal of Education/Revue Canadienne De l'éducation*, [online] 34(2), pp.3–20. Available at: <https://journals.sfu.ca/cje/index.php/cje-rce/article/view/876>.

Caskey, M., Peterson, K. and Temple, J. (2001). Complex Admission Selection Procedures for a Graduate Preservice Teacher Education Program. *Teacher Education Quarterly*, [online] 28(4), pp.7-21. Available at: <https://www.jstor.org/stable/23478313>.

Chang, M., Shaeffer, S., Al-Samarrai, S., Ragatz, A., De Ree, J. and Stevenson, R. (2014). *Teacher Reform in Indonesia: The Role of Politics and Evidence in Policy Making*. Directions in Development. [online] Washington, D.C.: World Bank. Available at: <https://openknowledge.worldbank.org/handle/10986/16355>.

Chen, D. (2009). *The Economics of Teacher Supply in Indonesia*. World Bank Policy Research Working Paper No. 4975. [online] Available at: <https://ssrn.com/abstract=1427635>.

Cruz-Aguayo, Y. and Ibarrarán, P. (2017). *Do Tests Applied to Teachers Predict Their Effectiveness?*. IDB-WP-821. [online] Inter-American Development Bank. Available at: <https://publications.iadb.org/en/do-tests-applied-teachers-predict-their-effectiveness>.

Davies, P., Connolly, M., Nelson, J., Hulme, M., Kirkman, J. and Greenway, C. (2016). 'Letting the Right One In': Provider Contexts for Recruitment to Initial Teacher Education in the United Kingdom. *Teaching and Teacher Education*, [online] 60, pp.291-302. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0742051X16303365#!>.

Goldhaber, D. (2019). Evidence-Based Teacher Preparation: Policy Context and What We Know. *Journal of Teacher Education*, [online] 70(2), pp.90-101. Available at: <https://journals.sagepub.com/doi/abs/10.1177/0022487118800712?journalCode=jtea>.

Goldhaber, D., Grout, C. and Huntington-Klein, N. (2017). Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools. *Education Finance and Policy*, [online] 12(2), pp.197-223. Available at: <https://direct.mit.edu/edfp/article-abstract/12/2/197/10270/Screen-Twice-Cut-Once-Assessing-the-Predictive?redirectedFrom=PDF>.

Hanushek, E., Lavy, V. and Hitomi, K. (2008). Do Students Care about School Quality? Determinants of Dropout Behavior in Developing Countries. *Journal of Human Capital*, [online] 2(1), pp.69-105. Available at: <https://www.journals.uchicago.edu/doi/abs/10.1086/529446?journalCode=jhc>.

Harris, D. and Sass, T. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics*, [online] 95(7-8), pp.798-812. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0047272710001696>.

Hill, H., Umland, K., Litke, E. and Kapitula, L. (2012). Teacher Quality and Quality Teaching: Examining the Relationship of a Teacher Assessment to Practice. *American Journal of Education*, [online] 118(4), pp.489-519. Available at: <https://www.journals.uchicago.edu/doi/abs/10.1086/666380>.

Huang, A., Revina, S., Fillaili, R. and Akhmadi. (2020). *The Struggle to Recruit Good Teachers in Indonesia: Institutional and Social Dysfunctions*. RISE Working Paper 20/041. [online] RISE Programme. Available at: <https://rise.smeru.or.id/en/publication/struggle-recruit-good-teachers-indonesia-institutional-and-social-dysfunctions>.

Hwa, Y. and Pritchett, L. (2021). *Teacher Careers in Education Systems That Are Coherent for Learning: Choose and Curate Toward Commitment to Capable and Committed Teachers (5Cs)*. [online] Research on Improving Systems of Education (RISE). Available at: <https://riseprogramme.org/publications/teacher-careers-education-systems-are-coherent-learning-choose-and-curate-toward>.

Jacob, B., Rockoff, J., Taylor, E., Lindy, B. and Rosen, R. (2018). Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools. *Journal of Public Economics*, [online] 166, pp.81-97. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0047272718301555>.

Kane, T., Rockoff, J. and Staiger, D. (2008). What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, [online] 27(6), pp.615-631. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0272775707000775>.

Kementerian Riset, Teknologi, dan Pendidikan Tinggi. (2017a). *Panduan PPG Pendidikan Profesi Guru 2017 [PPG Teacher Professional Education Guide 2017]*. Jakarta: Direktorat Pembelajaran Ditjen Belmawa Kementerian Riset, Teknologi, dan Pendidikan Tinggi.

Kementerian Riset, Teknologi, dan Pendidikan Tinggi. (2017b). *Prosedur Operasional Baku (POB) Seleksi Calon Mahasiswa Program PPG Prajabatan Bersubsidi 2017 [Standard Operating Procedure of the Selection of Prospective Students for the 2017 Subsidised Pre-Service PPG Programme]*. Direktorat Pembelajaran Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi, dan Pendidikan Tinggi.

Klassen, R. and Kim, L. (2019). Selecting Teachers and Prospective Teachers: A Meta-Analysis. *Educational Research Review*, [online] 26, pp.32-51. Available at: <https://www.sciencedirect.com/science/article/pii/S1747938X18302173>.

Lee, Y. and Kinzie, M. (2011). Teacher Question and Student Response with Regard to Cognition and Language Use. *Instructional Science*, [online] 40, pp.857-874. Available at: <https://link.springer.com/article/10.1007/s11251-011-9193-2>.

Malvern, D. (1991). Assessing the Personal Qualities of Applicants to Teacher Training. *Studies in Educational Evaluation*, 17, pp.239-253.

Metzler, J. and Woessmann, L. (2012). The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation. *Journal of Development Economics*, [online] 99(2), pp.486-496. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0304387812000478>.

Ragatz, A. (2015). *Indonesia - A Video Study of Teaching Practices in TIMSS Eighth Grade Mathematics Classrooms: Understanding What Teaching Practices are Used, Why They are Used and How They Relate to Student Learning*. Main Report. [online] Washington, D.C.: World Bank. Available at: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/886911472471847117/indonesia-a-video-study-of-teaching-practices-in-timss-eighth-grade-mathematics-classrooms-understanding-what-teaching-practices-are-used-why-they-are-used-and-how-they-relate-to-student-learning>.

Rarasati, N., Dharmawan, G., Swarnata, A., Zulfa, A. and Lim, D. (2020). *Comprehensive Reading and Mathematics Assessment Tool (CERMAT)*. Technical Report. [online] Jakarta: The SMERU Research Institute. Available at: <https://smeru.or.id/en/publication/comprehensive-reading-and-mathematics-assessment-tool-cermat>.

Rivkin, S., Hanushek, E. and Kain, J. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, [online] 73(2), pp.417-458. Available at: <https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2005.00584.x>.

Rockoff, J., Jacob, B., Kane, T. and Staiger, D. (2011). Can You Recognize an Effective Teacher When You Recruit One?. *Education Finance and Policy*, [online] 6(1), pp.43-74. Available at: <https://direct.mit.edu/edfp/article-abstract/6/1/43/10196/Can-You-Recognize-an-Effective-Teacher-When-You?redirectedFrom=fulltext#.VTfrStzF9yw>.

Rosser, A. (2018). *Beyond Access: Making Indonesia's Education System Work*. Analyses. [online] Lowy Institute. Available at: <https://www.lowyinstitute.org/publications/beyond-access-making-indonesia-s-education-system-work>.

Siraj, I. and Taggart, B. (2014). *Exploring Effective Pedagogy in Primary Schools: Evidence from Research*. [ebook] London: Pearson. Available at: <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/ExploringEffectivePedagogy.pdf> [Accessed 16 October 2018].

Staiger, D. and Rockoff, J. (2010). Searching for Effective Teachers with Imperfect Information. *Journal of Economic Perspectives*, [online] 24(3), pp.97-118. Available at: <https://www.aeaweb.org/articles?id=10.1257/jep.24.3.97>.

Vagi, R., Pivovarova, M. and Miedel Barnard, W. (2017). Keeping Our Best? A Survival Analysis Examining a Measure of Preservice Teacher Quality and Teacher Attrition. *Journal of Teacher Education*, [online] 70(2), pp.115-127. Available at: <https://journals.sagepub.com/doi/abs/10.1177/0022487117725025?journalCode=jtea>.

Wang, W., Dong, Y., Liu, X., Bai, Y. and Zhang, L. (2020). The Effect of Parents' Education on the Academic and Non-Cognitive Outcomes of Their Children: Evidence from China. *Children and Youth Services Review*, [online] 117, p.105307. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0190740920312093>.

World Bank. (2008). *Teacher Employment and Deployment in Indonesia: Opportunities for Equity, Efficiency and Quality Improvement*. Report. [online] Washington, D.C.: World Bank. Available at: <https://openknowledge.worldbank.org/handle/10986/8110> [Accessed 9 August 2020].

World Bank. (2019). *Ending Learning Poverty: What Will It Take?.* Report. [online] Washington, D.C.: World Bank. Available at: < https://openknowledge.worldbank.org/handle/10986/32553>.

Yusrina, A. and Bima, L. (2020). *Classroom Observation Tool for Assessing the Dimensions of Teaching Practices (CERDAS)*. Technical Report. [online] Jakarta: The SMERU Research Institute. Available at: <https://smeru.or.id/en/publication/classroom-observation-tool-assessing-dimensions-teaching-practices-cerdas>.

Appendix A

Table A.1. Distribution of Admission Criteria and Exit Exam Scores Data of 1,291 Teachers

	mean	min	p25	p50	p75	max
Admission criteria						
Undergraduate GPA	3.64	3.25	3.51	3.67	3.75	3.90
Overall Online Admission Score	56.76	51.59	55.37	56.55	58.63	64.60
- Score on Academic Aptitude Test	69.66	53.33	66.67	68.89	73.33	84.44
- Score on English Test	43.47	22.50	37.50	42.50	47.50	62.50
- Score on Pedagogical Knowledge Test	44.11	27.14	37.14	42.86	48.57	84.29
Interview Score	90.54	78.33	86.67	92.50	95.00	97.50
Exit exam scores						
Knowledge exam Score	82.33	75.25	79.67	82.32	83.21	90.28
Teaching practice exam Score	91.10	82.95	87.72	92.13	94.94	97.13

Appendix B

The figures below show that the correlation between the online admission test score and the interview and the correlation between the knowledge exam score and the teaching practice exam score are small (pairwise correlation coefficient of -0.06 and 0.07, respectively). The dashed line shows a linear fit. This means that the online admission test and the interview, and the knowledge and performance exam, might measure different things.

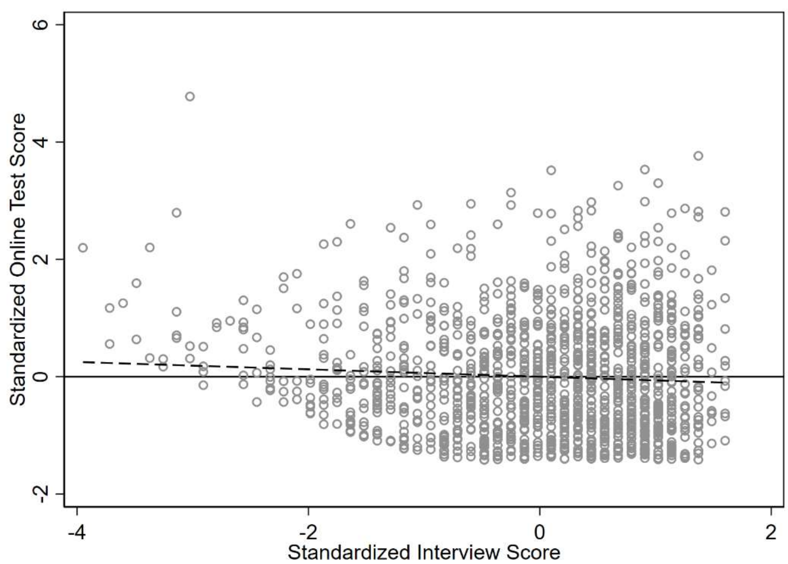

Figure A.1 Relationship between Screening Test Scores

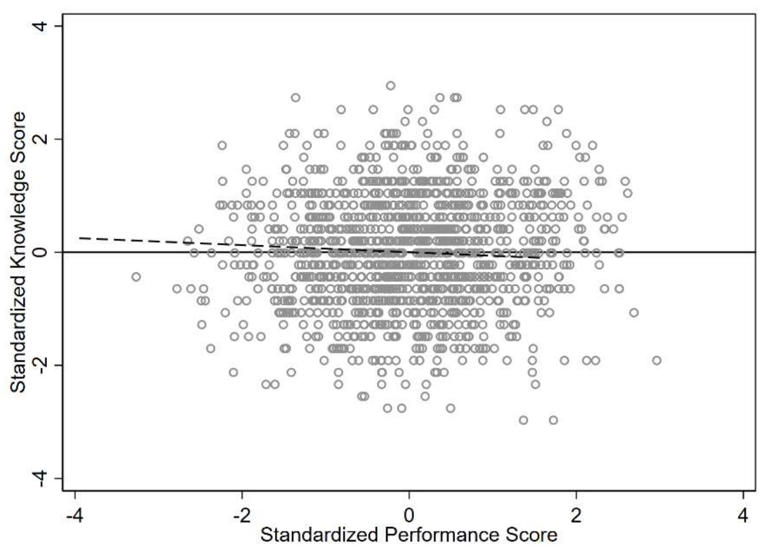

Figure A.2 Relationship between Exit Examination Scores

Appendix C

Student Learning Assessment

We developed a student learning assessment (SLA) instrument of numeracy and literacy, CERMAT, based on TIMSS' and PIRL s' (Progress in International Reading Literacy Study) frameworks, respectively, to assess the cognitive dimension of student's numeracy and literacy abilities. The numeracy assessment tool covers three primary mathematics content domains (numbers, geometry and measurement, as well as data and statistics) with the complexity of mathematical skill levels (from simple skill such as comparing numbers to a more advanced skill like algebraic operations). Meanwhile, the literacy assessment tool focuses on the reading skill levels adapted from The F&P Text Level Gradient™.

The composition of the items in the assessment tool is based on the results of multiple pilots conducted by the RISE Indonesia team. The numeracy test for lower-level primary school emphasises the knowing and applying abilities. Meanwhile, half of the test items for the higher-level primary school cover the knowing domain, while the rest are divided into the applying and reasoning. In the literacy assessment tool, the test assesses lower-level primary student's reading comprehension in several cognitive-process domains: early literacy, focus on and retrieve explicitly stated information, making straight-forward inferences, and interpret and integrate ideas and information. Furthermore, test items that assess the ability to evaluate and criticise content and textual elements are added to test higher-level primary students. Numeracy test questions are presented in either multiple-choice or closed constructed-response items. Meanwhile, the literacy test is in the form of multiple-choice, closed constructed-response items, and open constructed-response items.

The tests for lower-lever primary students are administered in one-on-one setting in which a trained enumerator will read the questions and ask the student to answer them. Meanwhile, higher-level students will do the test in a classical setting. In assessing the results, we use Item Response Theory (IRT) to calculate each student's proficiency by taking into account the number of questions answered correctly and their respective difficulty level.

Appendix D

Grade	Mean difficulties level of each grade		Number of teachers	Number of students	Score in Numeracy		Score in Literacy		Student's housing quality index
	Numeracy	Literacy			Mean	Std. Dev.	Mean	Std. Dev.	Mean
1	-2.93	-3.06	11	154	0	1	0	1	0.34
2	-2.45	-2.69	10	136	0.14	0.80	0.21	0.88	0.37
3	-1.68	-2.55	17	235	-0.07	0.83	-0.19	0.89	-0.13
4	0.62	-0.34	22	310	0.02	0.89	0.10	0.98	0.10
5	0.86	-0.12	31	439	0.11	1.05	-0.13	0.95	0.01
6	0.88	-0.66	23	346	0.25	1.00	0.21	1.00	0.12
			114	1620					