RISE Working Paper 23/134
March 2023

Screening Teachers in Indonesia: Does Ex-Ante Teacher Characteristics Assessment Predict Teaching Effectiveness?

Luhur Bima, Arjuni Rahmi Barasa, Shintia Revina, Niken Rarasati, and Asri Yusrina

Abstract

Selecting good teachers is vital as it can lead to a pool of teachers who will continuously strive to improve their teaching quality. Therefore, strengthening the assessment tools for screening effective teachers at their point of entry into the profession is important to improving teaching quality. While abundant studies have been conducted on improving teacher screening strategies in developed countries, only few have examined the process in the contexts of developing countries. Our study aims to enrich the literature on improving teacher screening in developing countries by testing sixty-two teachers using a set of teacher assessment instruments that measure both cognitive and non-cognitive skills. We discovered a significant and positive correlation between teacher competence in numeracy and student numeracy achievement. Furthermore, assessing teaching practices using a lesson demonstration is positively associated with students' achievement. However, we found a significant but negative relationship between teacher competence in literacy and student literacy outcomes. We also reported a similar pattern in the correlation between teachers' portfolio assessments and students' learning outcomes. The negative correlation in literacy measurements may be explained by the difficulties experienced by teachers in Indonesia in translating their knowledge into practice, as there are no specific subjects designated to Indonesian language and reading comprehension. From a policy perspective, the government and education institutions can strengthen their teacher recruitment mechanisms by adopting instruments that can predict teacher effectiveness. Furthermore, these screening instruments should be combined with ex-post assessment tools as those assessments will provide a comprehensive overview of teacher capabilities, not only in terms of prospective teacher characteristics but also in terms of their actual classroom teaching performance after a certain period of teaching practice.

Keywords: teacher screening instrument, teacher effectiveness

UKaid from the British people
Australian Aid
Bill & Melinda Gates foundation
Oxford Policy Management
Blavatnik School of Government
University of Oxford

Screening Teachers in Indonesia: Does Ex-Ante Teacher Characteristics Assessment Predict Teaching Effectiveness?

Luhur Bima
The SMERU Research Institute

Arjuni Rahmi Barasa
The SMERU Research Institute

Shintia Revina
The SMERU Research Institute

Niken Rarasati
The SMERU Research Institute

Asri Yusrina
The SMERU Research Institute

Acknowledgements:

We are grateful to Sudarno Sumarto, Menno Pradhan, and Daniel Suryadarma for their review of this paper.

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

Please cite this paper as:
Bima, L., Barasa, A.R., Revina, S., Rarasati, N. and Yusrina, A. 2023. Screening Teachers in Indonesia: Does Ex-Ante Teacher Characteristics Assessment Predict Teaching Effectiveness? RISE Working Paper Series. 23/134. https://doi.org/10.35489/BSG-RISE-WP_2023/134

Contents

List of Tables

1. Introduction

Teachers play an essential role in influencing student outcomes, not only in the short term, for example in learning outcomes as reported by Crawfurd and Rolleston (2020), Canales, A. and Maldonado, L. (2018), and Metzler and Woessmann (2012), but also in the long term, for example in earnings in adulthood (Chetty et al., 2014). Nevertheless, improving teacher quality to support a better learning environment for students is onerous. For decades, cross-national research has sought to examine how to improve teacher productivity (Bruns et al., 2018; Taylor and Tyler, 2012; Darling-Hammond, 2010; Hyslop-Margison and Sears, 2010; Nengwekhulu, 2008). However, there is an evident absence of a silver bullet that works in all contexts.

A straightforward option for improving teacher performance is to offer better financial incentives. However, the results of such mechanisms remain unclear, especially in the context of developing countries (Imberman, 2015). For example, using school datasets in the United States, Figlio and Kenny (2007) found that performance-based teacher financial incentives were positively associated with higher student test scores. On the other hand, an evaluation of the Indonesian teacher certification allowance program by De Ree et al. (2018) reported that financial incentives improved teachers' satisfaction but not students' learning outcomes. Similarly, Wang et al. (2014) found that a teacher performance-based pay program in China did not improve education quality. The program's restrictions and regulations caused teachers to focus on fulfilling the indicators of performance pay rather than enhancing their teaching skills.

Literature suggests that selecting high-quality teachers is vital, as it can lead to a pool of teachers who will continuously strive to improve their teaching quality. Kini and Podolsky (2016) state that the benefits of teaching experience to improving teacher expertise are better targeted at carefully selected teachers. Moreover, based on their analysis of the Florida Education Department dataset, Chingos and Peterson (2011) argue that while development training for in-service teachers may have little impact on its teachers' quality, state certification examination was more likely to identify effective teachers. Therefore, strengthening the mechanisms for screening teachers at their point of entry into the teaching profession is critical to improving teacher quality.

For years, scholars have attempted to improve screening strategies to effectively predict the future performance of prospective teacher candidates (e.g., Klassen et al., 2020; Kane et al., 2008). Cruz-Aguayo et al. (2017) evaluated the predictability of teacher candidate screening instruments used in public schools in the United States, the Urban Teacher Perceiver Interview and the Academic Skill Assessment portion of the Praxis Test on teacher performance measurement. However, their study did not link screening scores with student learning outcomes as an indicator of teacher performance. In another study, Rockoff and Speroni (2011) conducted an in-depth teacher survey to collect information on various teacher aspects, including cognitive ability, content knowledge (CK), personal traits, and personal beliefs and values. Their analysis of the relationship between these factors and student test scores suggested that a broad set of teacher measurements can enhance select teacher quality rather than focusing on individual aspects separately.

While abundant studies have been conducted on improving teacher screening strategies in developed countries, only few have examined the process in the contexts of different developing countries. Several of the few studies examine teacher recruitment reform in Latin American nations. For example, Cruz-Aguayo et al. (2017) found no evidence that teacher evaluation processes used in Ecuador—including a written test, a demonstration class, and other measurements—could predict teacher effectiveness in improving student learning outcomes. Meanwhile, Bertoni et al. (2020) evaluated the teacher hiring system in Peru, which consists of two stages: a centralized and decentralized selection processes. The authors reported mixed results in terms of the correlation between teacher screening tests and teacher performance measurement.

In screening and selecting effective teachers, Staiger and Rockoff (2010) proposed an ex-post evaluation in which teaching practices are assessed in a real classroom during a probationary period. High-performing teachers are then retained and can continue to teach. In line with the study by Staiger and Rockoff, Hwa and Pritchett (2021) emphasized the importance of teacher performance assessment at the end of the probationary period in order to determine teachers' true ability to perform in the classroom.

Although the ex-post evaluation method could be more effective at selecting the correct candidates as it provides more information on their teaching performance, in practice, this approach can prove challenging due to several constraints. One primary limitation is that the method requires more frequent hiring, as new teacher candidates must be recruited to replace teachers who fail to pass the probationary period. However, institutions responsible for recruiting teachers (both schools and governments) may experience the cost of teacher turnover, including its negative effect on students (Zeitilin, 2021; DeFeo et al., 2017; Watlington et al., 2010). Additionally, the system must be transparent and accountable to ensure that all stakeholders accept the assessment. Otherwise, frictions may arise due to unfair selection processes (Khanal, 2011).

This study aims to contribute to the discussion on effective teacher selection by evaluating the assessment tools for observable teacher characteristics, for example cognitive skills, in the context of developing countries. As the literature on teacher screening implementation remains relatively limited in developing countries, the current study adds further empirical evidence on whether teacher screening activities are significantly correlated with student learning outcomes. Moreover, the study evaluates whether teacher screening tools are sufficient to effectively predict the performance of prospective teacher candidates or whether these tools must be substituted or complemented with other approaches in order to optimize the search for effective teachers who can help students to improve their learning outcomes.

The analysis of our study is similar to the findings of Rockoff and Speroni (2011), yet the nature of our study differs in that we have the opportunity to test our set of teacher assessment instruments in a high-stakes environment. We collaborated with a local government in Indonesia that wished to assess sixty-two school-hired contract teachers (commonly known as honorary[1] teachers) in public schools under its jurisdiction to gain information on the teacher quality so that the government can improve the education quality in its region. Unlike data collected through surveys or taken from a low-stakes environment, which may suffer from downward bias in teacher measurement results, the current study's setting allowed us to ensure that the participants in the teacher assessment were highly motivated to take the tests seriously. As a result, the assessment results should reflect participants' abilities more accurately.

We use primary data on teachers and students for our analysis. The sample for the study is the population of school-hired contract teachers in public primary schools in one district in Indonesia. The data includes sixty-two teacher respondents who were assessed using various test instruments. We measured the teachers' cognitive skills in terms of pedagogical content knowledge (PCK) and CK—for both literacy and numeracy subjects—and numerical professional knowledge. In addition to these measurements, we also assessed the teachers' teaching practices by conducting interviews in order to determine their perceptions of how they implement teaching activities, to evaluate their teaching demonstrations, and to examine their portfolios.

[1] Honorary teachers are teachers hired by public schools to address teacher supply shortages. These teachers are informally hired (without a formal contract) and are largely underpaid.

We also assessed the learning of students taught by the teacher respondents in literacy and numeracy subjects, thus allowing us to estimate the association between teacher skills and student achievement. Around 1,400 students participated in the two-round student learning assessment, which was conducted before and after the teachers were assessed. The Least Absolute Shrinkage and Selection Operator (LASSO) regression method was employed to analyze the data, primarily to estimate the relationship between teacher characteristics measured by our instruments and student achievements (Ranstam and Cook, 2018; Tibshirani, 1996). This method helped us to select relevant controlled covariates for the model from a large number of available potential variables in order to obtain optimum regression results.

We found that teacher competence in numeracy is significantly associated with student numeracy outcomes. The composite of our three numeracy assessment instruments is positively correlated with student numeracy test scores, where one standard deviation higher of this composite is associated with an additional 0.11 standard deviation in the students' average scores. Furthermore, if the teacher numeracy measurements are analyzed separately, only their competence measurement in professional numeracy is significantly correlated with the students' score. An increase of one standard deviation in teacher numerical professional score is associated with an additional 0.10 standard deviation in student numeracy outcomes.

While the relationships between our measurements on teacher numeracy competence and student achievement are significant and positive, we learnt that the teachers' competence in literacy PCK tends to correlate negatively with student literacy achievements. An increase in the teachers' literacy PCK score by one standard deviation is associated with a 0.07 point lower standard deviation in the students' scores. If the students are divided into low- and high-ability students based on their initial abilities, the teachers' numeracy skills are positively correlated with student achievement in both groups. Nevertheless, this pattern does not occur in literacy. We found that only students in the low-ability group were influenced by teacher competence. Meanwhile, there is no evidence of a relationship between teacher literacy and student outcomes in the high-ability group.

The negative correlation between the teacher's literacy PCK and student's literacy scores must be interpreted cautiously. Teachers may have difficulty translating their knowledge into practice, as there are no specific subjects designated to Indonesian language and reading comprehension learning. Moreover, the materials presented in Indonesian language lessons suggested by the national curriculum do not focus on developing students' reading abilities, especially at the higher cognitive level of reading. Instead, Indonesian language lessons focus largely on vocabulary, grammar, and knowledge of various text genres (Minister of Education and Culture Regulation No. 37/2018).

We also uncovered significant relationships between several measurements of teachers' non-cognitive skills and student numeracy scores. Assessing teaching practice using a lesson demonstration is positively associated (0.08 standard deviation) with student achievement. On the other hand, we found that the teachers' portfolios are negatively correlated (-0.08 standard deviation) with the students' numeracy scores. Furthermore, our analysis shows that none of the non-cognitive measurements have significant relationships with student learning outcomes in literacy.

From a policy perspective, the government and education institution can strengthen their teacher recruitment mechanisms by adopting instruments that can predict teacher effectiveness, for example teaching demonstration assessment and numeracy skills assessment, when screening prospective teacher candidates. Furthermore, the assessment of prospective teachers should be complemented by an ex-post assessment after a probationary period. This assessment will provide

information on the candidates' actual capabilities to teach in a classroom, which cannot be observed in the early screening phase of the assessment (Hwa and Pritchett, 2021).

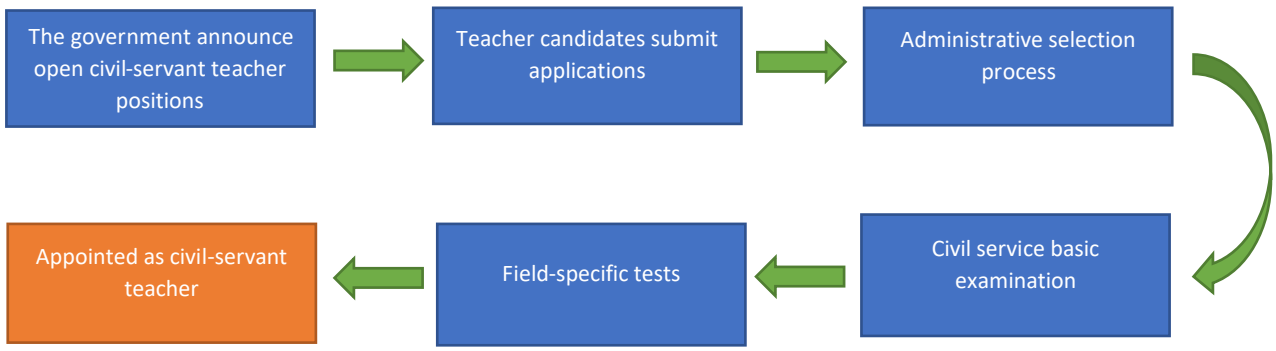
The next section briefly discusses how teacher competence is assessed in Indonesia. Section 3 provides a description of the data and method used in the current study. We present the results of our analysis in Section 4 and its discussion in Section 5. We conclude and provide recommendations in Section 6.

2. The Teacher Hiring Process in Indonesian Public Schools

In Indonesia, the teacher workforce is dominated by civil servant teachers (52%)[2] recruited by the government to teach in public schools. Being a civil servant teacher is highly demanded since it offers various benefits. Civil servant teacher in Indonesia receives teacher certification allowance in addition to the basic wage and other civil servant's allowances. Furthermore, it has a high job security level since the possibility of being dismissed from the position is very low. These advantages attract a huge pool of candidates who apply for this position. As the applicants exceed the available positions, the government conducts a selection mechanism in the hiring process. In the past, the selection process was based only on administrative formality. Nevertheless, as the bureaucratic system reform has progressed, the civil servant teacher candidates must pass the enrollment test to win the position (Huang et al, 2020).

Civil servants, including teachers, are recruited using the national civil service exam. Information and arrangements for application processes for teaching positions are handled by the district government (for early childhood to junior secondary school) and the provincial government (for senior secondary school). Meanwhile, the examination tools and system are developed and organized centrally by the Ministry of Administrative and Bureaucratic Reform.

Figure 1. Civil-Servant Teacher Selection Procedure



As shown in Figure 1, prospective teacher candidates who wish to become a civil-servant teacher must complete several steps in the recruitment process. Applicants for teaching positions are required to indicate one specific position at the school of their choice. This selection cannot be changed once they have submitted their application. Applicants for teaching positions who meet the

[2] According to data from the Ministry of Education and Culture, there are approximately 2,906,000 teachers in Indonesia. 52% of these are civil servants and 24% are public school-honorary teachers. Moreover, the rest of teacher workforce are in the private schools.

administrative requirements are eligible to take the civil service exam, which consists of the same basic competency tests as those taken by applicants for other vacant public sector positions (i.e., nationalism, psychology, and general intelligence tests), with an identical passing grade. In each test round, the top three candidates with the highest scores for each vacant position will be shortlisted for the next round, where they will take field-specific tests according to their subject. The shortlisted teacher candidates will take pedagogical knowledge and content knowledge tests. Teaching demonstrations and interviews are not required as part of the public teacher recruitment process. The candidate with the highest score on the test for each vacant teaching position will be automatically appointed as a civil servant teacher. These teachers are virtually on track for life tenure within the civil service (Huang et al., 2020).

Although the central government regulation formally prohibits public schools from recruiting non-civil servant teachers, many public schools and local governments continue to hire contract teachers to address teacher shortages that the central government cannot resolve due to budget constraints. Contract teachers are usually novice teachers who have just graduated from the teacher preparation programme and are awaiting the opening of civil servant teacher positions. Many contract teachers who fail the civil servant exam continue working informally in schools for many years. It is important to note that teachers fail the civil servant exams due to a number of reasons, including administrative requirements or moderate scores in the general basic competence test relative to other applicants. Unsuccessful attempts at passing the civil servant exam do not necessarily reflect teachers' lack of pedagogical or professional knowledge and skills.

Unlike civil servant teachers who are required to pass several tests to attain their positions, honorary teachers are recruited informally by schools. If any, the selection process of these teachers often only requires them to meet administrative criteria, for example holding a bachelor's degree certificate. As the central government does not formally recruit contract teachers, their salary is usually covered by the school's operational fund, which is extremely limited (Usman et al., 2007). This is one of the main reasons why school principals who take the initiative to recruit contract teachers do not apply any assessments to teacher candidates and are willing to accept contract teachers with minimal qualifications. Nevertheless, in several districts with greater budgets, district governments are willing to finance the salary of contract teachers by employing candidates as district-hired contract teachers.

The current study attempts to improve the existing civil-servant teacher recruitment process by modifying the field-specific test component. The default test only consists of pedagogical knowledge and content knowledge written tests. Based on previous literature, we propose adding several additional instruments that are positively correlated with teaching effectiveness and learning outcomes, including teaching practice assessments and interviews. Our hypothesis is that, if the modified field-specific tests have a significant impact on learning outcomes, they can improve the overall recruitment procedure of civil-servant teachers. Additionally, the modified assessment tools may be useful for schools that take the initiative to hire their own teachers in order to fill gaps produced by teacher retirement or resignation that cannot be immediately addressed by the central government.

3. Methodology

Our main study sample consists of contract-based primary school teachers and their students in one city in Indonesia. We collected information on the teachers' cognitive and non-cognitive skills through a set of assessments that we developed. We also assessed the literacy and numeracy learning outcomes of the students taught by the sample teachers. Having information on the

teachers' skills and their students' outcomes allows us to analyze which types of observable teacher skills may be significantly associated with student outcomes. We employed a one-step, lagged value-added model in performing our analysis. Moreover, the LASSO regression was implemented to optimize the regression analysis by selecting a set of control covariates that have non-zero coefficients.

3.1. Context

In this study, we focus on contract teachers recruited by public primary schools in Bukittinggi, a city in West Sumatera Province. Like other districts in Indonesia, the Bukittinggi City Government allows public schools in its area to recruit contract teachers to address teacher shortages. In 2019, the Bukittinggi Government wished to improve the welfare of its honorary teachers by promoting them to district-hired contract teachers[3]. Teachers with the latter status receive salaries at the regional minimum wage standard, i.e., around five times the salary of honorary teachers. Additionally, Bukitinggi, among other districts in Indonesia, has been known as a district with strong educational culture. The local community acknowledges education as a very important aspect. Parents prioritize their children's education so that their children can obtain higher education achievements to increase their social status. Thus, readers are supposed not to view Bukittinggi as a representative of districts in Indonesia.

Before promoting these contract teachers, the Bukittinggi Government wished to examine the teachers' levels of competence and what needed to be improved to enhance educational outcomes in Bukittinggi. We took advantage of this opportunity by assisting the Government to assess the contract teachers' competence. The collaboration allowed us to conduct the assessment in a high stake testing environment, as the Bukittinggi Government had agreed to inform the teachers that the assessment results would affect their promotion. For the assessment, we developed a set of tools to gauge the teachers' competencies, including cognitive and non-cognitive skills. When developing the tools, we considered their predictive ability to identify effective teachers and their ease of use for school staff members and local governments responsible for selecting and hiring teachers.

As part of the collaboration, we informed the Bukittinggi Government of the results of the teacher assessments. The teacher participants were ranked based on their assessment performance. Furthermore, the Bukittinggi Government decided, based on various factors including the political interest of the mayor of Bukittinggi, that all teacher participants would be contracted by the Bukittinggi Government. Moreover, the Bukittingi education agency used the assessment results to encourage schools to enhance their teachers' skills. The Bukittinggi Government also planned to use the results as input for developing teacher training programs.

Approximately sixty-two classroom teachers participated in our assessment. We also assessed around 1,400 students taught by the teachers participating in the assessment. The purpose of examining the students' learning outcomes in numeracy and literacy was to identify which teacher selection instruments were associated with student learning outcomes. In addition, we also surveyed the students' parents to identify family characteristics that can potentially affect student outcomes.

With a permission from the Bukitinggi government, we collected student data in two rounds. The first was conducted in August 2019, two months before the teacher assessment. Furthermore, the second round of student data collection was conducted in October 2020, which was the third month

[3] District-hired contract teachers are teachers who have a formal contract from the district government and paid by the government from the district budget.

of the new academic year for 2020/2021. Our initial plan was to complete the assessments before the 2019/2020 academic year ended. However, the pandemic forced us to postpone the activities. The second round of data collection suffered from student attrition in which around 239 students who participated in the first round of data collection were not able to participate in the second round of assessments due to various reasons. Nevertheless, as presented in Table A.4 in the appendix, we found that the mean scores between the students who completed both assessments and those who did not take the second round of testing for numeracy are not statistically different. On average, there is an attrition rate of four student for each sample teacher. Despite there being three teachers with more than ten students who only attended the first round of assessments, we found that the mean scores between sample obs1 and sample obs2 for numeracy are not statistically different. However, student literacy scores could be affected by student attrition, despite the size being relatively small.

It is important to note that learning activities were significantly affected by the pandemic situation during our second round of data collection. The sample students in the first round testing cohort were required to study from home from March to October 2020. During the second round of data collection, they had been taught by new teachers for approximately three months. We are aware that the teachers in the following year level may have affected the students' learning outcomes, as we conducted the second assessment in the subsequent academic year. Nevertheless, the attribution of the new teachers during this short period of time is expected to be relatively small, as learning activities remained suboptimal due to the learning from home policy. The second round of data collection was required to be conducted by visiting students in their homes, as the seven schools were not yet open for learning activities. Despite expecting that any systematic bias due to the pandemic would be minimal as all student samples equally experienced the situation, it is important to note that the sample teachers were not able to teach optimally, which could lead to the underestimation of the contribution of teachers' skills to student learning outcomes.

The research sample for the current study focuses on contract teachers in one city (Bukittinggi City). Consequently, the relevance of our results may be limited only to this group of teachers and may not be representative of teachers in general in Indonesia. The characteristics of other types of teachers, for example civil servant teachers, may be different from our sample. Nevertheless, the results of the current study are still insightful in understanding the correlation between teacher characteristics and student learning outcomes.

3.2. Teacher and Student Assessment Instruments

Research on teacher recruitment focuses on analyzing the teacher factors that matter to student learning outcomes and what screening instruments can effectively predict the performance of prospective teacher candidates. Previous studies have suggested that teacher cognitive skills positively impact student achievement. For instance, Tchoshanov et al. (2017) found that, in lower secondary schools, the correlation between the mathematical CK acquired by a teacher and student performance in mathematics is statistically significant. Baumert et al. (2010) also reported a significant effect of a teacher's PCK on student achievement through cognitive activation and individual learning support channels.

Furthermore, other studies have argued that teacher non-cognitive skills can also positively influence student achievement. Kane et al. (2011) analyzed the ability of teaching practices to predict student achievement based on the Cincinnati Teacher Evaluation System (TES) dataset, which evaluates teaching practices in four domains: (i) Planning and Preparing for Student Learning, (ii) Creating an Environment for Student Learning, (iii) Teaching for Student Learning, and (iv) Professionalism. The authors found that a higher ability to manage the classroom environment may

lead to higher student learning outcomes. Similarly, Allen et al. (2011) suggested that the quality of student-teacher interactions may predict student learning gains.

We developed screening tools to identify teachers' cognitive and non-cognitive skills and correlate the skills with student learning outcomes. Table 1 describes our screening tools, their objectives, and the aspects that each tool attempts to measure.

Table 1 Description of Screening Tools

Tool	Objective	Measure
Teaching portfolio document	This is a preliminary assessment to determine the teachers' capacity in terms of teaching methods, performing student assessments, developing their skills, and teaching evaluations.	We adopted the teaching portfolio assessment used in developed countries, including the United Kingdom and Australia. Teachers were requested to submit materials that reflect their teaching development progress, including examples of their teaching tools, examinations or quizzes that measure children's learning progress, proof of training or workshops attended related to teaching development, etc. We developed a rubric that can be used by independent evaluators based on the following aspects: (i) teaching methods, (ii) student assessment, (iii) development of (teaching) skills, and (iv) teaching evaluation.
Content knowledge (CK) test (written test consisting of fifteen questions)	To examine the levels of teachers' CK in literacy and numeracy.	To determine the link between teacher subject mastery and student subject knowledge in numeracy and literacy, we selected approximately 30% of items from the CERMAT (Comprehensive Reading and Mathematics Assessment) (Rarasati et al., 2020).
Pedagogical content knowledge (PCK) test (written test consisting of fifteen questions)	To examine the levels of teachers' PCK in literacy and numeracy.	We assessed two components in the PCK test: knowledge of teaching concepts (KTC) and knowledge of students' knowledge (KSK) (Olfos et al., 2014). In the KTC component, we assessed teachers' constructivist orientation toward learning and their knowledge of organizing their teaching. Meanwhile, in the KSK component, we evaluated teachers' knowledge on student learning, for example common errors made by students in specific topics and their thoughts on these topics.
Professional numeracy (PN) skills (written test consisting of ten questions)	To measure basic literacy and numeracy skills, regardless of the teachers' expertise.	The test measures teachers' ability to use data and research to diagnose student learning processes (Guerriero, 2017). This ability helps teachers to reflect, evaluate, and improve their teaching practices. We adopted the professional skills test administered by the United Kingdom's Department of Education.
Teaching practice test (15-minute teaching demonstration session)	To assess teachers' communication and classroom management skills, as well as their concept and understanding of certain	We developed a rubric that can be used by independent evaluators to measure general pedagogical knowledge and professional skills (e.g., communication and instructional decision-making skills). The rubric measures five aspects: (i) introduction (opening activities), (ii) presentation, (iii)

	subjects in daily teaching activities.	interaction with students, (iv) teaching aids, and (v) attitude (poise and professional conduct). The assessment rubric for teaching demonstrations can be seen in Table A.5 in the appendix.
Interview (15-minute session)	To assess teachers' communication skills, and attitudes and beliefs towards teaching and student learning.	We developed a rubric that evaluates teachers' answers to questions about differentiation of teaching based on students' competence, student assessment, and their experience in doing both. The assessment rubric for teacher interviews can be seen in Table A.6 in the appendix.

For PCK, we designed questions that incorporate the same curricula (i.e., geometry, reading comprehension) as the CK tests to determine how teachers deliver their subject mastery to the students. We used items response theory (IRT) for all written tests to create a score on teachers' latent ability based on items with different difficulty levels.

For the teaching practice test and interview, we were interested in assessing teachers' communication skills. Teachers demonstrate communication skills at school, particularly with colleagues, students, and parents; thus, such skills should be assessed. Cook and Friend (1993) argued that fostering collaboration with peers provides many benefits in dealing with instructional issues. Teachers with the ability to communicate well in a team are more likely to discuss the curriculum and other difficulties, resulting in a more robust instruction plan in the classroom. Regarding teacher-student interaction, Frymier and Houser (2000) identified that interpersonal communication between teachers and students leads to more effective teaching, as it strongly relates to student learning and motivation. Furthermore, effective communication with parents may help teachers improve student achievement as this provides a deeper understanding and a mutual expectation of students' needs (Sheldon and Epstein, 2005).

Student Achievement Test

We assessed the students' numeracy and literacy abilities using the CERMAT (Comprehensive Reading and Mathematics Assessment) tool (Rarasati et al., 2020). CERMAT is a literacy and numeracy assessment tool developed by the RISE Indonesia team based on the Indonesian 2006 and 2013 curriculum combined with TIMSS and PIRLS numeracy and literacy cognitive domain frameworks. The instrument was developed to support studies conducted by the RISE Indonesia team, as there are no existing standardized literacy and numeracy tests that could be used for our analysis. Hence, this instrument is mainly used in studies conducted by the RISE Indonesia team.

The instrument was designed to assess the skills of students in grades 1 to 9 in understanding mathematical concepts, applying ideas to solve real-life problems, analyzing, synthesizing, evaluating, and performing other types of systematic thinking using mathematical concepts. For reading, the assessment covers skills such as retrieving explicit information, making inferences, integrating information, interpreting implicit information, and applying critical thinking to reading materials.

The current study uses booklets for grades 1 to 6 all items in which are written in Indonesian. To capture the students' learning progress while being taught by the sample teachers, we conducted two rounds of CERMAT tests. The first round was conducted in August 2019 in school settings. 1,803 students taught by 63 classroom teachers participated in the test. The second round was carried out during the COVID-19 pandemic in October 2020. Hence, we made several adjustments in

administering the test. First, the test was performed at students' homes, monitored by enumerators. Second, we cut the test length by about 35% (from twenty-one to fourteen items on average for numeracy and sixteen to eleven items on average for literacy). Attrition occurred during the second round of the test. Students who were in grade 6 in the first round had already graduated by the time we held the second round of testing. Moreover, due to the pandemic's health and safety concerns, not all parents were willing to have their homes visited for testing. The attrition left 1,414 students of 62 classroom teachers who completed the second-round of testing.

We conducted two rounds of data collection to examine how the students progressed after being taught by the sample teachers for one year. We compared the two test scores fairly by employing a two-step equating method. The first step was implemented during the test design. CERMAT booklets were designed to have sufficient anchor items (35%) between grades and between the two test rounds (Rarasati et al., 2020). The second step was to obtain students' latent ability scores based on items' parameter estimates based on IRT analysis for all grades and test rounds. This step enabled the score to have the same scale across grades and test rounds (Cook and Eignor, 1991). We re-examined the effectiveness of the anchoring method in the equating process using IRT and found that several items appeared to have incongruence difficulty levels when comparing the estimate based on psychometric analysis and qualitative judgment according to the content domain. We dropped seven numeracy items and four reading comprehension items from the analysis to minimize bias estimation due to the incongruence difficulty level.

Teacher Content Knowledge Testing

To capture the link between teacher subject mastery and student subject knowledge in numeracy and literacy, we selected approximately 30% of items from CERMAT for the CK test. The test was designed for students in grades 1 to 9.

According to IRT analysis, we used items from third to sixth-grade booklets with higher difficulty levels according to item response theory (IRT) analysis based on data taken during the CERMAT development process (Rarasati et al., 2020). For PCK tests, we designed problems that incorporate the same curricula (i.e., geometry, reading comprehension) as CK tests to assess how teachers delivered their subject mastery to the students.

3.3. Analytical Method

The current study identifies which aspects of teacher attributes correlate with student learning outcomes. Our analysis uses the one-step lagged model that has been widely used to estimate teacher effectiveness in other studies (Koedel et al., 2015). We estimate the correlation between teacher competencies and student learning value-added using the following regression model:

$$Y_{ikt} = \alpha_0 + Y_{ikt-1} + \beta \cdot \boldsymbol{S}_k + \gamma_1 \cdot \boldsymbol{X}_{ik} + \gamma_2 \cdot \boldsymbol{V}_k + \gamma_3 \cdot \boldsymbol{Z} \quad + \varepsilon_i$$

Where Y is the learning outcome in literacy or numeracy of student *i* taught by teacher *k* in period *t.* Meanwhile, **S** is a vector of teacher-written test instruments, portfolio, interview score, and microteaching score, **X** is a vector of student characteristics including grade, average class size and test score, and parents' education. **V** is a teacher background vector consisting of age, teaching experience, and education, while **Z** is a school-level variable. E is an error term.

Our dataset consists of various student, parent, and teacher variables. On one hand, this allowed us to explore appropriate control variables relevant to our model. However, on the other hand, having

too many potential covariates could cause an overfitting issue in the estimation. In our dataset, around sixty-five variables could be considered as potential covariates. Hence, we used the LASSO regression method to help select control variables in our estimation. The LASSO regression is a machine learning technique that minimizes prediction error, providing only a limited number of non-zero coefficients (Ranstam and Cook, 2018; Tibshirani, 1996). For instance, McEligot et al. (2020) applied logistic LASSO regression to identify and select the most relevant variables associated with self-reported breast cancer status. Additionally, we used the Stata 17 software in analyzing the data using the LASSO regression approach.

4. Results

4.1. Descriptive statistics

We began our analysis by examining the teacher participants' performance in the assessment. As presented in Table A.2 in the appendix, we determined that from the measurements of teacher competence in numeracy, teachers performed better in professional numeracy, as it represents the basic knowledge that should be acquired by all teachers to evaluate the progress of student learning. Furthermore, we uncovered that teachers assigned to higher grades tended to achieve higher scores in all written tests. We found that the variation of teacher skills was quite large, in which teachers assigned to grade two were more likely to perform badly compared to teachers of other grades. Teachers tended to demonstrate better competence in non-cognitive skills than in cognitive skills. Moreover, we found that the variation in teachers' non-cognitive skills across grades was relatively small.

Table 2 presents the pairwise correlation of student test scores and teacher assessment results. We can see that, at the student level, the 2020 literacy scores are positively associated with the 2020 numeracy scores. We note that students' 2020 literacy scores have no significant correlation with teachers' literacy CK or PCK. On the other hand, the correlations between teachers' numeracy skill measurements are substantial. The students' 2020 numeracy scores significantly correlate with teachers' PCK, both in literacy and numeracy. Furthermore, although the correlation is minimal, we see that student literacy and numeracy scores are positively correlated with teachers' teaching demonstration assessments.

Overall, most teacher tests are significantly correlated, although the size varies across test instruments. We note that significant correlations appear in teachers' numeracy-related skill measurements. The correlation between teacher numeracy knowledge and numeracy pedagogical knowledge is approximately 0.75. Meanwhile, the correlation between numeracy knowledge and numerical professional skill is around 0.62. Although the size is relatively small, there is also a significant correlation between literacy CK and literacy PCK. In addition to this, the assessment scores on microteaching seem to have a significant and positive correlation with the other tests, except for PCK in literacy.

Table 2 Pairwise Correlation of Student and Teacher Scores

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) SLIT19	1											
(2) SLIT20	0.3452*	1										
(3) SNUM19	0.3117*	0.2517*	1									
(4) SNUM20	0.2586*	0.3421*	0.2949*	1								
(5) CKB	-0.0186	0.0320	-0.0021	0.0138	1							
(6) PCKB	-0.0532*	-0.0489	-0.0754*	-0.0556*	0.308*	1						
(7) CKM	-0.0613*	0.0609*	0.0293	0.0353	0.2277*	0.1987*	1					
(8) PCKM	-0.0184	0.0664*	0.0038	0.0491*	0.0620*	0.1752*	0.7466*	1				
(9) NP	-0.0537*	-0.0195	0.003	0.0233	0.1210*	0.1852*	0.6222*	0.5369*	1			
(10) INT	0.0054	0.0143	0.0264	-0.0212	0.0835*	0.1563*	-0.0466	-0.003	-0.1071*	1		
(11) TEACH	0.0047	0.0837*	-0.0496*	0.0818*	0.0867*	0.0150	0.1572*	0.1904*	0.1478*	-0.0840*	1	
(12) PORT	-0.0242	-0.0081	-0.03	-0.024	-0.1647*	-0.2158*	0.0331	0.0574*	0.0746*	-0.0960*	0.2241*	1

* $p < 0.10$

(1) SLIT19 : Student literacy score 2019
(2) SLIT20 : Student literacy score 2020
(3) SNUM19 : Student numeracy score 2019
(4) SNUM20 : Student numeracy score 2020
(5) CKB : Content knowledge – literacy
(6) PCKB : Pedagogical content knowledge – literacy
(7) CKM : Content knowledge – numeracy
(8) PCKM : Pedagogical content knowledge – numeracy
(9) NP : Professional numeracy
(10) INT : Interview
(11) TEACH : Teaching demonstration
(12) PORT : Portfolio

4.2. Investigation of potential biases

Before proceeding to the main model analysis, we will examine the correlation between the characteristics of our sample teachers and their respective students. We will attempt to identify whether the assignment of our teachers to classes was random or not. Non-randomized teacher assignments may potentially create a bias in our estimation, as teachers with higher initial qualifications could be purposively assigned to specific grades, for example grade 6, where students require intensive preparation for final exams.

We regressed the student's grades in relation to teacher characteristics, including age, undergraduate GPA, years of teaching experience, gender, type of university, and their study program. Our sample is dominated by female teachers. Furthermore, on average, the teachers had six years of teaching experience. Additionally, most of our sample (80%) graduated from the School Teacher Education Program (See Table A.1 in the Appendix for the description).

The results of the estimation are presented in Table 3. The table shows that only the type of university from which the teachers graduated is significantly correlated with teaching assignment to grade. This finding shows that teachers who graduated from public universities are more likely to be assigned to teach higher grades. Nevertheless, we found that this is the only teacher characteristic that may affect the randomness of teacher assignment to grade. The other teacher characteristics seem to be not significantly correlated with teaching assignment. We attempted to minimize the non-randomized teacher assignment due to the teacher's university background by controlling the teacher's characteristics.

Table 3 Regression Results between Teacher Characteristics and ***C****lassroom* ***G****rade*

	(1) grade
Teacher is female	-0.62 (0.83)
Teacher's age	0.04 (0.04)
Teacher graduated from public university	1.05* (0.60)
Teacher graduated from primary school teacher education program	-0.08 (0.67)
Teacher's undergraduate GPA	0.03 (0.74)
Years of teaching experience	-0.06 (0.06)
Observations	1336

Standard errors in parentheses and clustered at teacher level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We are aware that the teachers were not randomly assigned to their schools, as schools can hire teachers based on their preferences. Furthermore, the students' learning achievement may also have been influenced by their background, for example their parent's education. As a result, the estimation results may have been influenced by the potential biases that may have occurred due to the characteristics of the students and teachers. We attempted to minimize this bias by controlling variables at the student and teacher levels.

4.3. Estimation results

Table 4 estimates the correlation between the teacher performance assessment instruments and student numeracy outcomes. We constructed a teacher numeracy index as a composite of teacher numeracy CK, PCK scores, and numerical professional scores. The estimations of the teacher numeracy index in relation to the students' numeracy scores can be seen in Column 1 (without control variables) and Column 3 (with control variables). We see that the correlation between the numeracy index and student numeracy scores is insignificant without additional covariates in the model. However, the correlation became statistically significant once we controlled several selected variables. One standard deviation higher in the teacher numeracy index is associated with a 0.11 standard deviation in student numeracy scores.

Furthermore, we found that the positive correlation between teacher demonstration scores and student numeracy scores is significant after the control variables are included. We also noticed an unexpected negative and significant correlation between teacher portfolio and student numeracy scores. One standard deviation higher in teacher portfolio is associated with a 0.08 standard deviation lower in student numeracy scores.

We estimated the correlation between teacher numeracy assessment instruments and student numeracy achievement. The teacher numeracy CK measurement was excluded due to its high correlation with numeracy PCK. When the control variables are not included in the estimation model (Column 2), the numerical PCK of the teacher is significantly associated with student numeracy scores. Raising the numerical PCK score by one standard deviation is associated with a 0.09 standard deviation increase in student numeracy scores.

However, this correlation changes once control variables are added. We found that, once control variables are added, the numerical PCK becomes insignificant (Column 4). Meanwhile, a significant and positive correlation between teacher numerical professional skill and student numeracy score was reported. An increase of one standard deviation in teacher numerical professional score is associated with an additional 0.10 standard deviation in student numeracy outcomes. We also found the exact relationship between teaching demonstrations and student achievement, although the coefficient is slightly lower (0.08 standard deviation).

The correlation between teacher portfolio and numeracy scores is negative and significant. A teaching portfolio is documentation that reflects the teacher's individual professional development journey (Weeks, 2006), which implies that teachers with more extensive portfolios tend to be more experienced and skillful. However, it can be argued that a teaching portfolio may inaccurately reflect actual teaching ability and skill, which is important for student learning (Whitworth et al., 2016). This argument implies that teachers with a better portfolio are not guaranteed to be more effective teachers. Furthermore, there may also be a trade-off between preparing teaching materials for student learning and improving their portfolio if teachers do not clearly understand what matters in producing student learning outcomes.

Table 4 Regression Results between Teacher Assessment Components and Student Numeracy Achievement

	(1)	(2)	(3)	(4)
Teacher Numeracy Index	0.07*		0.11***	
	(0.04)		(0.03)	
Numeracy Pedagogical Content Knowledge Score		0.09**		0.04
		(0.04)		(0.04)
Numeracy Professional Score		-0.02		0.10***
		(0.04)		(0.04)
Teaching Demo Score	0.07	0.06	0.08**	0.08***
	(0.04)	(0.04)	(0.03)	(0.03)
Interview Score	0.02	0.01	0.03	0.04
	(0.04)	(0.04)	(0.03)	(0.03)
Portfolio Score	-0.04	-0.04	-0.08**	-0.08**
	(0.05)	(0.05)	(0.03)	(0.03)
Student 2019 Numeracy Test Score	0.35***	0.35***	0.35***	0.35***
	(0.03)	(0.03)	(0.03)	(0.03)
Student Grade	-0.14***	-0.15***	-0.16***	-0.16***
	(0.03)	(0.03)	(0.03)	(0.03)
Constant	0.38***	0.39***		
	(0.09)	(0.09)		
Observations	1403	1403	1106	1106
r2	0.12	0.13	-	-
Available controls	-	-	66	66
Selected controls	-	-	17	19

Standard errors in parentheses and clustered at teacher level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We calculated the teacher literacy index as a composite of literacy content and PCK scores and estimated it with other teacher assessment instruments in relation to student literacy achievement (Table 5). As we can see in Columns 1 and 3 of Table 5, the teacher literacy index has no significant correlation with student outcomes, regardless of the control variables. Furthermore, we found that teaching demonstration score is significantly associated with student literacy scores when control variables are excluded from the estimation. However, the correlation becomes insignificant once we control additional characteristics.

The estimation results in Column 4 show that teacher PCK in literacy becomes significantly associated with student outcomes after control variables are included in the regression. Nevertheless, the correlation is surprisingly negative, which means that one standard deviation higher in teacher competence in literacy PCK is associated with a 0.07 standard deviation lower in student scores.

Table 5 Regression Results between Teacher Assessment Components and Student Literacy Achievement

	(1)	(2)	(3)	(4)
Teacher Literacy Index	-0.01		-0.04	
	(0.05)		(0.04)	
Literacy Content Knowledge Test Score		0.04		-0.01
		(0.04)		(0.03)
Literacy Pedagogical Content Knowledge Test Score		-0.05		-0.07**
		(0.04)		(0.03)
Teaching Demo Score	0.07**	0.07**	0.03	0.04
	(0.03)	(0.03)	(0.03)	(0.03)
Interview Score	0.02	0.02	-0.00	0.02
	(0.05)	(0.05)	(0.03)	(0.03)
Portfolio Score	-0.02	-0.02	-0.04	-0.05
	(0.04)	(0.04)	(0.03)	(0.03)
Student 2019 Literacy Test Score	0.35***	0.34***	0.32***	0.33***
	(0.03)	(0.03)	(0.03)	(0.03)
Student Grade	-0.01	-0.01	0.04	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Constant	0.01	0.03		
	(0.09)	(0.10)		
Observations	1403	1403	1106	1106
r2	0.12	0.13	-	-
Available controls	-	-	66	66
Selected controls	-	-	18	18

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4. Estimation Results for Low vs High Ability Students

In this section, we evaluate whether the effect of teachers may be different across students. We divided the students into two groups: low and high ability. Students who's initial (2019) test scores were below the median are categorized as low-ability students whereas the high-ability group scored above the cut-off. We can see in Columns 1 and 3 of Table 6 that the teachers' numeracy index is significantly associated with low-ability students' performance, whether the other covariates are excluded (0.10 standard deviation) or included (0.13 standard deviation). Meanwhile, the numeracy index has a significant correlation with the numeracy outcome of high-ability students only after we add the control variables into the model specification. Column 7 of Table 6 shows that an additional standard deviation in the teachers' numeracy index is associated with a 0.10 standard deviation higher in student scores.

When we added the teacher's numeracy measurements as separate indicators into a simple estimation where no control variable was included, we found that the correlation between

numeracy PCK and low-ability student numeracy scores is significant (0.11 standard deviation), but not for the numerical professional score (Column 2). Once we included the control covariates, the positive correlation of numeracy PCK remains significant (0.10 standard deviation), while the numerical professional score is now also positively and significantly (0.09 standard deviation) associated with the numeracy performance of low-ability students (Column 4).

The results for the high-ability group are slightly different. The relationship between numeracy PCK and student performance was significant before controlling the other covariates. Nevertheless, the correlation changes to be insignificant once we include the control variables. On the other hand, the correlation between numerical professional score and student score changed from negative and insignificant to positive and significant after the control variables were added.

The estimations also show that the teacher's teaching skill positively correlates with students' numeracy scores in both the low- and high-ability groups. However, the magnitude is slightly smaller in the high-ability group. An increase in the teaching demonstration score of one standard deviation is associated with a 0.12-0.13 standard deviation increase in the low-ability students' scores and a 0.07-0.08 standard deviation increase in the opposite student group.

Additionally, in the previous section, we discussed the negative correlation between a teacher's portfolio and students' numeracy scores. Here we note that the teacher portfolio scores have a negative and significant correlation with low-ability students' scores, but not with high-ability students' scores.

Table 7 provides the estimation results of teacher assessment instruments and students' literacy outcomes. We see that teachers' literacy skill measurements, either as a composite index or individual score, have a significant and negative correlation with low-ability students' literacy scores. In Column 4, we see that one standard deviation higher in the teacher's pedagogical knowledge score is associated with a 0.11 standard deviation lower in the literacy score. Nevertheless, these significant correlations do not appear in high-ability students, although this group's correlation between teacher PCK and student outcome remains negative.

We discovered that a teacher's teaching practice may be important for low-ability students but not for high-ability students in relation to literacy skills. The estimations show that an increase of one standard deviation in teaching demonstration correlates with an improvement of 0.08-0.1 standard deviation in students' literacy outcomes. Additionally, we did not find any evidence of significant relationships between teacher interview and portfolio scores and students' literacy performance for low- and high-ability students.

Table 6 Regression Results between Teacher Assessment Components and Student Numeracy Achievement—Low vs High Ability Students

	Low Ability Students				High Ability Students			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher Numeracy Index	0.10**		0.13***		0.05		0.10**	
	(0.04)		(0.05)		(0.05)		(0.05)	
Numeracy Pedagogical Content Knowledge Test Score		0.11**		0.10*		0.08*		0.02
		(0.05)		(0.06)		(0.05)		(0.06)
Numerical Professional Test Score		0.00		0.09*		-0.03		0.11*
		(0.04)		(0.06)		(0.05)		(0.05)
Teaching Demo Score	0.06	0.06	0.12***	0.13***	0.05	0.04	0.07*	0.08*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Interview Score	0.04	0.04	0.05	0.05	-0.03	-0.04	0.01	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
Portfolio Score	-0.08	-0.08	-0.12***	-0.12***	-0.00	-0.00	-0.01	-0.01
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Student 2019 Numeracy Test Score	0.23***	0.24***	0.24***	0.26***	0.29***	0.29***	0.28***	0.26***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.07)	(0.07)
Student Grade	-0.17***	-0.17***	-0.16***	-0.17***	-0.10**	-0.10***	-0.13***	-0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Housing Quality Index	0.01	0.01	-0.00	0.01	0.00	0.01	-0.01	-0.02
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Father Completed University Degree	0.08	0.09	-0.07	-0.06	0.29**	0.28**	0.28**	0.30**
	(0.12)	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)
Mother Completed University Degree	0.26***	0.25***	0.33***	0.33***	0.11	0.11	0.12	0.12
	(0.09)	(0.09)	(0.13)	(0.13)	(0.09)	(0.09)	(0.11)	(0.11)

Constant	0.26**	0.27**			0.26*	0.27*		
	(0.12)	(0.13)			(0.14)	(0.14)		
Observations	700	700	548	548	703	703	558	558
r2	0.08	0.09			0.07	0.07		
Available controls	-	-	63	63	-	-	63	63
Selected controls	-	-	36	31	-	-	26	25

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7 Regression Results between Teacher Assessment Components and Student Literacy Achievement—Low vs High Ability Students

	Low Ability Students				High Ability Students			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher Literacy Index	-0.03		-0.08		0.02		-0.01	
	(0.07)		(0.05)		(0.04)		(0.05)	
Literacy Content Knowledge Score		0.06		-0.04		0.04		0.02
		(0.05)		(0.05)		(0.04)		(0.04)
Literacy Pedagogical Content Knowledge Score		-0.09		-0.09*		-0.01		-0.02
		(0.06)		(0.05)		(0.04)		(0.05)
Teaching Demo Score	0.06	0.05	0.09**	0.10**	0.05	0.05	0.04	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)
Interview Score	0.03	0.04	0.05	0.06	-0.01	-0.01	-0.04	-0.03
	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Portfolio Score	-0.04	-0.03	-0.03	-0.03	0.01	0.00	-0.02	-0.03
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)
Student 2019 Literacy Test Score	0.36***	0.37***	0.36***	0.38***	0.31***	0.31***	0.29***	0.29***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
Student Grade	-0.06	-0.07	0.03	0.04	0.06	0.05	0.09**	0.07*
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Housing Quality Index	-0.04	-0.05*	-0.05	-0.05*	0.02	0.02	0.01	0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Father Completed University Degree	0.36***	0.37***	0.36***	0.35***	0.09	0.09	0.11	0.10
	(0.12)	(0.12)	(0.12)	(0.12)	(0.10)	(0.10)	(0.10)	(0.10)
Mother Completed University Degree	0.07	0.08	0.17	0.13	0.17**	0.17**	0.12	0.15
	(0.10)	(0.09)	(0.12)	(0.12)	(0.08)	(0.08)	(0.10)	(0.10)

Constant	0.12 (0.13)	0.15 (0.13)			-0.19* (0.11)	-0.18 (0.11)		
Observations	729	729	570	570	674	674	536	536
r2	0.09	0.10	-	-	0.06	0.06		
Available controls	-	-	63	63	-	-	63	63
Selected controls	-	-	27	32	-	-	28	30

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5. Discussion

Based on the estimation results presented in the previous section, we found that teacher performance has a significant and positive association with students' numeracy scores, especially for low ability students. Nevertheless, we did not discover an exact significant correlation with literacy scores. Furthering our analysis, we learnt that students' numeracy skills were more likely to have a significant correlation with our teacher measurements than students' literacy skills.

The positive correlation between teaching demonstration scores and students' numeracy scores implies that it is crucial to assess a teacher's ability to translate their knowledge into teaching practices that lead to higher student achievement. This finding, to some extent, is similar to a study conducted by Bruno and Strunk (2019), which found that sample lesson assessment has the ability to predict a teacher's contribution to their students' achievement. Nevertheless, it is important to note that in their study, the correlation between the sample lesson assessment and students' math scores is relatively weak compared to students' English and literacy scores.

In contrast to the findings of Bruno and Strunk's (2019) study, the current study found an insignificant correlation between lesson practice assessment (teaching demonstrations) and students' literacy performance. Two possible factors may influence this finding. First, external factors may have a contribution to students' reading comprehension abilities than language lessons at school. Gumus and Atalmis (2011) found that students who spent a substantial amount of time using computer-based entertainment obtained a significantly higher literacy score on PISA compared to those who only used technology for school activities. Marx and Stanat (2012) also made a similar finding, in that access to various reading resources other than school textbooks played a significant role in student reading comprehension skills.

The second factor can be derived from the Indonesian national curriculum. Since 2013, teachers have been required to integrate all subjects using a thematic approach (Yulianti, 2015). Many teachers experience difficulties adopting this approach, as the subjects were taught separately in the previous curriculum. As a result, as mathematics is perceived to require a different learning approach from other subjects, the government decided to revise the 2013 curriculum in 2018, allowing teachers to teach mathematics separately from other subjects (Sulistyani and Deviana, 2019; Minister of Education and Culture Regulation No. 37/2018). Teachers are also equipped with better teaching numerical lesson materials and textbooks. For the upper grades of primary school, the Ministry of Education's Centre for Curriculum and Book Affairs also provides mathematics textbooks in addition to the general thematic textbooks containing materials from various subjects. This is not the case with Indonesian language or reading literacy. Not only are there no available textbooks to teach the subject independently, but the materials in the Indonesian language lessons suggested by the national curriculum do not focus on developing students' reading ability, especially the higher cognitive level of reading. Rather, the language lessons focus on vocabulary, grammar, and knowledge of various genres of texts (Minister of Education and Culture Regulation No. 37/2018).

Regarding students' Indonesian literacy outcomes, several results are puzzling. For example, higher teacher scores in the literacy PCK tests are associated with lower student literacy scores. The negative correlation between PCK and student literacy scores must be interpreted cautiously. Based on interviews conducted with several of the teachers who participated in this study, students experience difficulties in reading comprehension. Despite being able to read the sentences in the reading materials assigned to them, they cannot process the information in order to understand and interpret the materials. This situation is exacerbated by the fact that the teachers may experience

difficulties translating their knowledge into practice, as there is no specific subject designated to teaching Indonesian language or reading comprehension. With the integrated subject method and no specific set targets for students' reading skills in the curriculum, teachers focus more on teaching other subjects that are clearly related to the theme of the lesson. Hence, students may not be given enough time to engage in lengthy and various reading materials, other than textbooks, in order to develop their literacy skills.

Additionally, when we divide the students into groups based on their initial ability, we can see that students with a low ability are the ones who are significantly affected by the teacher's PCK in language, rather than the high-ability students. Control variables, including socioeconomic status and parents' educational background, may indicate that the home literacy environment plays a greater role in students' literacy development than teacher-related factors (Ho and Lau, 2018).

Our analysis also indicates that low-ability students are more greatly affected by the teacher-related factors measured in the current study. For instance, the teacher's teaching practice skills significantly influenced the numeracy and literacy achievement of low-ability students. Meanwhile, an exact significant correlation was not found in the high-ability group. This implies that our screening tools may be helpful to identify teachers who effectively improve the learning outcomes of students with an initial lower ability. However, students with a higher initial ability may not receive the same benefits as those with lower ability students. Additionally, we found that the variation of competencies across the sample teachers was quite high (See Table A.2 in the Appendix). This indicates a high range of teacher quality, which implies that students can be taught by either a very high competence teacher or a very low competence teacher. Thus, low-ability students may be more severely impacted if they are taught by a teacher with a low competence level.

6. Conclusions and Recommendations

The current study assessed the association between teacher screening instruments and student learning outcomes in numeracy and literacy. Overall, the present study demonstrates that, to some extent, an objective measure of teacher performance in selecting teachers can identify teachers who produce higher achievement among their students. This is in line with previous studies on teacher selection that promote teacher performance indicators to identify teacher characteristics aligned with effectiveness and assist officials in teacher screening and hiring processes (Goldhaber et al., 2017; Kane et al., 2013; Goldhaber et al., 2014; Rockoff and Speroni, 2011).

From a policy perspective, the government and other relevant institutions can use teaching demonstration assessments to screen teacher candidates. Additionally, an assessment to measure teachers' competence in numeracy can be introduced at an early stage. The cost of administering the assessment is expected to be low[4], however test can help identify prospective teachers who will be able to improve student learning outcomes.

Nevertheless, the weak association between the screening instruments and student learning outcomes suggests that the screening tools must be complemented with other approaches in order to optimize the selection process. Hwa and Pritchett (2021) suggested that the novice phase of the teaching career is critical in identifying a teacher's actual ability to perform in the classroom. Assessing novice teachers' performance during the probation period will be beneficial, as classroom

[4] Administering the assessment test is expected to be relatively cheap, as schools and district governments can use existing facilities to organize the assessment. Moreover, the test materials can be regularly updated at a low cost using the government or school budget.

teaching practices can differ from the theories that teacher candidates learn during their education. Candidates' higher performance in the screening tests does not guarantee that better actual teaching practices in the classroom are positively associated with student learning outcomes. A study conducted by Bau and Das (2020) provided empirical evidence that observed teacher characteristics, including cognitive skills and teaching experience, have a very small contribution to student learning outcomes.

Therefore, educational institutions, and the government particularly in the context of civil-servant teachers, should consider the application of ex-ante and ex-post teacher assessments during the recruitment process. This combination of assessments will provide a comprehensive overview of teacher capabilities, not only in terms of prospective teacher characteristics but also in terms of their actual classroom teaching performance. Given that more extensive and accurate information on teacher capabilities can be acquired through this process, the assessments will increase the probability of the educational institution or the government recruiting more effective teachers.

Additionally, the instruments can also be utilized to evaluate teacher effectiveness in general. Our research results also provide information on which teacher-related aspects affect student achievement. Thus, the government can build on teacher development programs according to the needs of students in order to improve the learning process. The findings suggest that teachers must improve their ability to translate their pedagogical knowledge into teaching practice, particularly in teaching literacy.

However, as our study only includes honorary teachers, further studies may need to compare the results of the current study to the performance of civil servant teachers. While Suryadarma et al. (2006) suggested that the performance of schools with more civil servant teachers is not higher than schools with more honorary teachers, it would be valuable to examine how different pools of teachers perform using the same selection instruments developed above. Note that the selection instruments outlined above consist of different components to the Indonesian civil servant teacher test.

References

Allen, J., Pianta, R., Gregory, A., Mikami, A. and Lun, J. (2011). An Interaction-Based Approach to Enhancing Secondary School Instruction and Student Achievement. *Science*, [online] 333(6045), pp.1034-1037. Available at: <https://www.science.org/doi/10.1126/science.1207998>.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. and Tsai, Y. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, [online] 47(1), pp.133-180. Available at: <https://journals.sagepub.com/doi/abs/10.3102/0002831209345157>.

Bau, N., Das, J. (2017). The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers. RISE Working Paper 17/016. [online] RISE Programme. Available at: < https://riseprogramme.org/sites/default/files/2020-10/RISE_WP-016_Bau-Das.pdf>

Bertoni, E., Elacqua, G., Méndez, C. and Santos, H. (2020). *Teacher Hiring Instruments and Teacher Value Added: Evidence from Peru*. IDB Working Paper Series, No. IDB-WP-01199. [online] Washington, D.C.: Inter-American Development Bank. Available at: <https://publications.iadb.org/en/teacher-hiring-instruments-and-teacher-value-added-evidence-peru>.

Bruno, P. and Strunk, K. (2019). Making the Cut: The Effectiveness of Teacher Screening and Hiring in the Los Angeles Unified School District. *Educational Evaluation and Policy Analysis*, [online] 41(4), pp.426-460. Available at: <https://journals.sagepub.com/doi/abs/10.3102/0162373719865561>.

Bruns, B., Costa, L. and Cunha, N. (2018). Through the Looking Glass: Can Classroom Observation and Coaching Improve Teacher Performance in Brazil?. *Economics of Education Review*, [online] 64, pp.214-250. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0272775717303576>.

Canales, A. and Maldonado, L. (2018). Teacher Quality and Student Achievement in Chile: Linking Teachers' Contribution and Observable Characteristics. *International Journal of Educational Development*, [online] 60, pp.33-50. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S073805931730281X>.

Chetty, R., Friedman, J. and Rockoff, J. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, [online] 104(9), pp.2633-2679. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2633>.

Chingos, M. and Peterson, P. (2011). It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness. *Economics of Education Review*, [online] 30(3), pp.449-465. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0272775710001755>.

Cook, L. and Eignor, D. (1991). IRT Equating Methods. *Educational Measurement: Issues and Practice*, [online] 10(3), pp.37-45. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.1991.tb00207.x>.

Cook, L. and Friend, M. (1993). "Educational Leadership for Teacher Collaboration" in *Program Leadership for Serving Students with Disabilities*, Virginia Tech. pp. 421-444.

Crawfurd, L. and Rolleston, C. (2020). Long-Run Effects of Teachers in Developing Countries. *Review of Development Economics*, [online] 24(4), pp.1279-1299. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/rode.12717>.

Cruz-Aguayo, Y., Ibarrarán, P. and Schady, N. (2017). Do Tests Applied to Teachers Predict Their Effectiveness?. *Economics Letters*, [online] 159, pp.108-111. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0165176517302604>.

Darling-Hammond, L. (2010). *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*. Report. [online] Washington, D.C.: Center for American Progress. Available at: <https://www.americanprogress.org/article/evaluating-teacher-effectiveness/>.

De Ree, J., Muralidharan, K., Pradhan, M. and Rogers, H. (2018). Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia. *The Quarterly Journal of Economics*, [online] 133(2), pp.993-1039. Available at: <https://academic.oup.com/qje/article-abstract/133/2/993/4622956?redirectedFrom=fulltext>.

DeFeo, D., Tran, T., Hirshberg, D., Cope, D. and Cravez, P. (2017). *The Cost of Teacher Turnover in Alaska*. Report. [online] Anchorage, Alaska: Center for Alaska Education Policy Research, University of Alaska Anchorage. Available at: <https://scholarworks.alaska.edu/handle/11122/7815>.

Figlio, D. and Kenny, L. (2007). Individual Teacher Incentives and Student Performance. *Journal of Public Economics*, [online] 91(5-6), pp.901-914. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S004727270600140X>.

Frymier, A. and Houser, M. (2000). The Teacher-Student Relationship as an Interpersonal Relationship. *Communication Education*, [online] 49(3), pp.207-219. Available at: <https://www.tandfonline.com/doi/abs/10.1080/03634520009379209>.

Goldhaber, D., Cowan, J. and Theobald, R. (2017). Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. *Journal of Teacher Education*, [online] 68(4), pp.377-393. Available at: <https://journals.sagepub.com/doi/abs/10.1177/0022487117702582>.

Goldhaber, D., Walch, J. and Gabele, B. (2014). Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments. *Statistics and Public Policy*, [online] 1(1), pp.28-39. Available at: <https://www.tandfonline.com/doi/full/10.1080/2330443X.2013.856169>.

Guerriero, S. (ed.) (2017). Pedagogical Knowledge and the Changing Nature of the Teaching Profession. Paris: OECD Publishing. Available at: <https://www.oecd.org/education/pedagogical-knowledge-and-the-changing-nature-of-the-teaching-profession-9789264270695-en.htm>.

Gumus, S. and Atalmis, E. (2011). Exploring the Relationship between Purpose of Computer Usage and Reading Skills of Turkish Students: Evidence from PISA 2006. *Turkish Online Journal of Educational Technology*, [online] 10(3), pp.129-140. Available at: <http://www.tojet.net/volumes/v10i3.pdf>.

Ho, E. and Lau, K. (2018). Reading Engagement and Reading Literacy Performance: Effective Policy and Practices at Home and in School. *Journal of Research in Reading*, [online] 41(4), pp.657-679. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9817.12246>.

Huang, A., Revina, S., Filaili, 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> [Accessed 8 June 2021].

Hyslop-Margison, E. and Sears, A. (2010). Enhancing Teacher Performance: The Role of Professional Autonomy. *Interchange*, [online] 41, pp.1-15. Available at: <https://link.springer.com/article/10.1007/s10780-010-9106-3?noAccess=true>.

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] RISE Programme. Available at: <https://riseprogramme.org/sites/default/files/2021-12/Teacher_careers_education_systems_coherent_learning.pdf>

Imberman, S. (2015). How Effective are Financial Incentives for Teachers?. *IZA World of Labor*, [online] Available at: <https://wol.iza.org/articles/how-effective-are-financial-incentives-for-teachers/long>.

Kane, T., McCaffrey, D., Miller, T. and Staiger, D. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Research Paper. [online] Seattle: Bill & Melinda Gates Foundation. Available at: <https://usprogram.gatesfoundation.org/news-and-insights/usp-resource-center/resources/have-we-identified-effective-teachers-validating-measures-of-effective-teaching-using-random-assignment>.

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>.

Khanal, P. (2011). Teacher Management in a Decentralised School Context in Nepal: Fuelling Tension and Dissent?. *Compare: A Journal of Comparative and International Education*, [online] 41(6), pp.769-784. Available at: <https://www.tandfonline.com/doi/abs/10.1080/03057925.2011.592041>.

Kini, T. and Podolsky, A. (2016). *Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research*. [online] Palo Alto: Learning Policy Institute. Available at: <https://learningpolicyinstitute.org/product/does-teaching-experience-increase-teacher-effectiveness-review-research>.

Klassen, R., Kim, L., Rushby, J. and Bardach, L. (2020). Can We Improve How We Screen Applicants for Initial Teacher Education?. *Teaching and Teacher Education*, [online] 87, p.102949. Available at: <https://www.sciencedirect.com/science/article/pii/S0742051X18322030>.

Koedel, C., Mihaly, K. and Rockoff, J. (2015). Value-Added Modeling: A Review. *Economics of Education Review*, [online] 47, pp.180-195. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0272775715000072>.

Ladd, H. and Sorensen, L. (2017). Returns to Teacher Experience: Student Achievement and Motivation in Middle School. *Education Finance and Policy*, [online] 12(2), pp.241-279. Available at: <https://direct.mit.edu/edfp/article-abstract/12/2/241/10265/Returns-to-Teacher-Experience-Student-Achievement?redirectedFrom=fulltext>.

Marx, A. and Stanat, P. (2012). Reading Comprehension of Immigrant Students in Germany: Research Evidence on Determinants and Target Points for Intervention. *Reading and Writing*, [online] 25, pp.1929-1945. Available at: <https://link.springer.com/article/10.1007/s11145-011-9307-x>.

McEligot, A., Poynor, V., Sharma, R. and Panangadan, A. (2020). Logistic LASSO Regression for Dietary Intakes and Breast Cancer. *Nutrients*, [online] 12(9), p.2652. Available at: <https://www.mdpi.com/2072-6643/12/9/2652>.

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>.

Minister of Education and Culture Regulation No. 37/2018 on the Amendments to Minister of Education and Culture Regulation No. 24/2016 on Core Competencies and Basic Competencies of Lessons in the 2013 Curriculum in Basic and Secondary Education.

Nengwekhulu, R. (2008). Reflections on Improving Teacher Performance. *Africa Education Review*, [online] 5(2), pp.338-350. Available at: <https://www.tandfonline.com/doi/abs/10.1080/18146620802450041>.

Olfos, R., Goldrine, T. and Estrella, S. (2014). Teachers' Pedagogical Content Knowledge and Its Relation with Students' Understanding. *Revista Brasileira de Educação*, [online] 19(59), pp.913-944. Available at: <https://www.scielo.br/j/rbedu/a/b7mNsqsg4JLLSTyrP4pctFP/?lang=en>.

Ranstam, J. and Cook, J. (2018). LASSO Regression. *British Journal of Surgery*, [online] 105(10), p.1348. Available at: <https://academic.oup.com/bjs/article/105/10/1348/6122951?login=true>.

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>.

Rockoff, J. and Speroni, C. (2011). Subjective and Objective Evaluations of Teacher Effectiveness: Evidence from New York City. *Labour Economics*, [online] 18(5), pp.687-696. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0927537111000315>.

Sheldon, S. and Epstein, J. (2005). Involvement Counts: Family and Community Partnerships and Mathematics Achievement. *The Journal of Educational Research*, [online] 98(4), pp.196-207. Available at: <https://www.tandfonline.com/doi/abs/10.3200/JOER.98.4.196-207>.

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>.

Sulistyani, N. and Deviana, T. (2019). Analisis Bahan Ajar Matematika Kelas V SD di Kota Malang [Analysis of Mathematics Teaching Materials for Grade 5 in Malang City]. *Jurnal Pemikiran dan Pengembangan Sekolah Dasar*, [online] 7(2), pp.133-141. Available at: < https://ejournal.umm.ac.id/index.php/jp2sd/article/view/9692>.

Suryadarma, D., Suryahadi, A., Sumarto, S. and Rogers, F. (2006). Improving Student Performance in Public Primary Schools in Developing Countries: Evidence from Indonesia. *Education Economics*, [online] 14(4), pp.401-429. Available at: <https://www.tandfonline.com/doi/abs/10.1080/09645290600854110?journalCode=cede20>.

Taylor, E. and Tyler, J. (2012). Can Teacher Evaluation Improve Teaching?. *Education Next*, [online] 12(4), pp.78-84. Available at: <https://www.educationnext.org/can-teacher-evaluation-improve-teaching/>.

Tchoshanov, M., Cruz, M., Huereca, K., Shakirova, K., Shakirova, L. and Ibragimova, E. (2017). Examination of Lower Secondary Mathematics Teachers' Content Knowledge and Its Connection to Students' Performance. *International Journal of Science and Mathematics Education*, [online] 15, pp.683-702. Available at: <https://link.springer.com/article/10.1007/s10763-015-9703-9>.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, [online] 58(1), pp.267-288. Available at: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x>.

Usman, S., Akhmadi and Suryadarma, D. (2007). Patterns of Teacher Absence in Public Primary Schools in Indonesia. *Asia Pacific Journal of Education*, [online] 27(2), pp.207-219. Available at: <https://www.tandfonline.com/doi/abs/10.1080/02188790701378826>.

Wang, L., Lai, M. and Lo, L. (2014). Teacher Professionalism under the Recent Reform of Performance Pay in Mainland China. *Prospects*, [online] 44, pp.429-443. Available at: <https://link.springer.com/article/10.1007/s11125-014-9315-0#citeas>.

Watlington, E., Shockley, R., Guglielmino, P. and Felsher, R. (2010). The High Cost of Leaving: An Analysis of the Cost of Teacher Turnover. *Journal of Education Finance*, [online] 36(1), pp.22-37. Available at: <https://www.jstor.org/stable/40704404>.

Whitworth, J., Deering, T., Jones, S. and Hardy, S. (2016). Hiring Quality Teachers: The Devil Is in the Details. *The Professional Educator*, [online] 40(1). Available at: <https://files.eric.ed.gov/fulltext/EJ1103361.pdf>.

Yulianti, K. (2015). The New Curriculum Implementation in Indonesia: A Study in Two Primary Schools. *International Journal about Parents in Education*, [online] 9(1), pp.157-168. Available at: <https://www.researchgate.net/publication/288824214_The_New_Curriculum_Implementation_in_Indonesia_A_Study_in_Two_Primary_Schools>.

Appendices

Table A.1: Descriptive Statistics of Teacher and Student Characteristics

Variables	Mean	SD	Observation
Teacher			

Female	0.90	0.30	62
Age	32.05	6.52	62
Years of experience at current school	6.18	5.23	62
GPA	3.23	038	62
Graduated from School Teacher Education Program	0.81	0.40	62
Graduated from public university	0.39	0.49	62
Class size	23.67	4.48	62
School quality index	-1.01	0.92	62
Student			
Female	0.48	0.50	1,403
Age	9.32	1.40	1,403
No grade repetition	0.90	0.31	1,403
Family Asset index	0.00	1.47	1,403
Father completed university degree	0.14	0.35	1,403
Mother completed university degree	0.25	.43	1,403

Table A.2: Teacher Test Scores by Grade

Test Scores	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number of observations	17		14		16		10		5	
Written test										
Content Knowledge – Literacy	-0.04	0.72	-0.32	1.51	-0.12	0.75	0.40	1.00	0.63	0.41
Content Knowledge – Numeracy	0.10	0.85	-0.86	0.95	0.20	0.88	0.23	0.87	1.00	0.80
Pedagogical Content Knowledge – Literacy	-0.11	0.68	-0.16	1.22	0.27	0.91	-0.31	1.37	0.60	0.34
Pedagogical Content Knowledge – Numeracy	-0.05	1.03	-0.50	1.02	0.21	0.92	-0.02	0.89	0.93	0.72
Professional Numeracy	0.15	0.91	-0.57	1.06	0.34	1.03	-0.07	0.94	0.12	0.80
Practice test										
Interview score	0.11	0.81	0.10	1.03	-0.27	0.97	0.21	0.86	-0.15	1.62
Teaching demo score	0.37	0.74	-0.24	0.96	-0.32	1.20	0.37	0.66	-0.30	1.42
Portfolio Score	0.28	0.95	-0.43	0.67	0.18	1.17	-0.11	1.17	-0.11	0.88

Table A.3: Change of Student Scores by Grade

Grade	Number of Teachers	Number of Students	Literacy Score			Numeracy Score			Overall Gain
			Baseline	Endline	Gain	Baseline	Endline	Gain	
1	17	342	-0.26	0.35	0.61***	-0.62	0.52	1.13***	0.87***
			1.10	0.87		1.05	0.90		
2	14	334	0.29	0.36	0.07	0.20	0.73	0.53***	0.30***
			0.93	0.90		1.06	0.99		
3	16	345	-0.16	0.35	0.51***	0.08	0.38	0.29***	0.40***
			0.91	0.73		0.77	0.83		
4	10	251	0.14	0.31	0.17*	0.39	0.81	0.21***	0.19***
			0.99	0.99		0.60	0.95		
5	5	131	0.11	0.19	0.08	0.14	0.31	0.17	0.13
			0.91	0.98		0.86	0.99		

Note: Standard deviations are in parentheses Gain is the difference between standardized scores of the baseline and endline survey. We tested the gain in means with a t-test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Attrition Check on Student Numeracy and Literacy Tests

As several students did not participate in the second round of learning assessment for various reasons, we checked whether the students' test scores were statistically different from the test scores of students who participated in both rounds of assessment.

Obs1 is the sample of students who were not included in the second round of assessment, and obs2 is the sample of students who completed both assessments. Overall, we found that the mean scores for sample obs1 and sample obs2 for numeracy are not statistically different. However, for the literacy test, the mean score of obs1 is slightly lower than the mean score of obs2 with the difference being statistically significant. Consequently, the student literacy scores may have been affected by student attrition, although the number is relatively small.

Teacher Code	Obs1	Obs2	Numeracy				Literacy			
			Mean1	Mean2	Dif	p-value	Mean1	Mean2	Dif	p-value
All Teachers	239	1403	-0.05	0.00	-0.05	0.48	-0.10	0.02	-0.12	0.08
Teacher 1	2	26	-0.83	-0.72	-0.11	0.84	-0.51	-0.07	-0.44	0.63
Teacher 2	8	18	-0.66	-0.64	-0.02	0.96	0.10	0.45	-0.35	0.36
Teacher 4	7	23	0.62	0.33	0.29	0.32	0.88	0.25	0.63	0.17
Teacher 5	14	22	-0.39	0.13	-0.51	0.09	0.22	0.36	-0.13	0.66
Teacher 6	2	25	0.10	-0.03	0.13	0.83	-0.12	0.18	-0.30	0.66
Teacher 7	3	25	-0.50	-0.26	-0.24	0.77	-0.12	-0.18	0.07	0.91
Teacher 8	4	24	0.61	0.83	-0.22	0.76	0.30	0.81	-0.51	0.30
Teacher 9	2	29	0.86	0.13	0.73	0.16	0.27	0.50	-0.23	0.73
Teacher 10	1	30	1.61	0.31	1.30	.	0.72	0.36	0.37	.
Teacher 13	3	20	0.05	-0.91	0.97	0.16	-0.19	-0.46	0.27	0.76
Teacher 14	4	28	0.15	0.80	-0.65	0.21	0.65	0.85	-0.20	0.61
Teacher 15	3	25	-0.10	0.24	-0.33	0.35	-0.27	0.23	-0.50	0.35
Teacher 16	1	32	-0.18	0.33	-0.51	.	0.84	0.71	0.13	.
Teacher 17	2	26	1.37	0.31	1.06	0.12	0.07	-0.08	0.15	0.79
Teacher 18	2	25	0.44	0.40	0.04	0.94	0.58	-0.26	0.84	0.28
Teacher 19	1	26	0.54	0.25	0.30	.	0.92	-0.28	1.20	.
Teacher 20	1	26	0.15	-0.16	0.31	.	0.12	-0.21	0.33	.
Teacher 21	3	28	0.46	0.45	0.01	0.98	0.88	0.45	0.42	0.30
Teacher 22	1	24	-0.83	-1.04	0.21	.	-0.92	0.01	-0.93	.
Teacher 23	2	22	-0.13	0.46	-0.59	0.38	0.67	0.21	0.47	0.54
Teacher 24	1	22	0.94	-0.10	1.04	.	1.42	-0.44	1.86	.
Teacher 25	1	15	-1.86	-1.24	-0.62	.	0.52	-0.40	0.92	.
Teacher 26	1	24	-2.02	0.15	-2.17	.	-1.70	0.08	-1.78	.
Teacher 27	2	23	0.77	-0.16	0.93	0.24	0.51	0.06	0.45	0.56

Teacher Code	Obs1	Obs2	Numeracy				Literacy			
			Mean1	Mean2	Dif	p-value	Mean1	Mean2	Dif	p-value
Teacher 28	3	22	-0.62	0.19	-0.81	0.23	0.71	0.10	0.61	0.32
Teacher 29	5	30	0.57	0.97	-0.41	0.39	0.76	0.50	0.26	0.50
Teacher 30	5	20	-0.72	-0.55	-0.17	0.65	-0.44	-0.08	-0.36	0.38
Teacher 31	2	24	0.49	-0.64	1.13	0.21	-0.54	-0.56	0.03	0.98
Teacher 33	5	16	0.44	-0.21	0.64	0.23	-0.65	-0.05	-0.61	0.18
Teacher 34	2	31	1.04	0.28	0.77	0.16	-0.19	0.22	-0.40	0.57
Teacher 36	3	19	0.13	-0.28	0.41	0.42	0.33	-0.31	0.63	0.27
Teacher 39	8	16	0.21	-0.04	0.25	0.51	-0.34	-0.50	0.16	0.58
Teacher 40	5	19	-0.14	-0.42	0.28	0.62	0.44	0.34	0.09	0.84
Teacher 41	3	21	1.00	0.71	0.30	0.52	-0.36	-0.30	-0.06	0.90
Teacher 42	8	19	0.60	0.51	0.09	0.78	0.22	-0.02	0.24	0.51
Teacher 43	1	25	0.06	0.17	-0.12	.	0.31	0.17	0.14	.
Teacher 44	3	24	0.27	-0.02	0.29	0.65	0.09	0.11	-0.02	0.97
Teacher 46	3	24	0.42	-0.14	0.56	0.19	0.01	-0.54	0.55	0.27
Teacher 47	3	29	-0.08	-0.15	0.07	0.86	-0.30	0.09	-0.40	0.41
Teacher 48	17	18	0.96	0.53	0.43	0.16	-0.47	-0.54	0.07	0.82
Teacher 49	2	25	-0.35	-0.32	-0.02	0.98	0.02	-0.05	0.07	0.93
Teacher 50	9	17	-0.22	0.38	-0.61	0.04	0.11	0.18	-0.07	0.88
Teacher 51	15	7	-0.76	-0.80	0.04	0.95	-0.41	-0.72	0.31	0.53
Teacher 52	10	15	-0.36	-0.80	0.44	0.21	-0.51	-0.05	-0.45	0.24
Teacher 53	3	12	-0.23	0.12	-0.35	0.46	-0.82	-0.41	-0.41	0.51
Teacher 54	5	20	-0.33	0.21	-0.54	0.20	-0.26	0.21	-0.46	0.27
Teacher 55	4	14	-0.55	-1.07	0.52	0.53	-2.24	-0.14	-2.10	0.01
Teacher 56	3	17	-0.39	-0.76	0.38	0.49	-0.02	-0.25	0.22	0.70
Teacher 57	8	20	-0.77	-0.49	-0.28	0.53	-0.68	-0.75	0.08	0.88
Teacher 58	2	24	-0.27	0.30	-0.56	0.49	-0.48	0.13	-0.61	0.39
Teacher 59	8	20	0.39	0.09	0.30	0.32	0.35	0.30	0.05	0.87
Teacher 60	5	21	-0.02	0.21	-0.23	0.60	-0.40	0.00	-0.40	0.40
Teacher 61	4	17	0.89	0.57	0.32	0.44	0.14	0.07	0.06	0.89
Teacher 62	3	21	-1.51	-0.57	-0.94	0.05	-0.66	-0.34	-0.32	0.62
Teacher 63	1	25	0.94	-0.12	1.05	.	-0.33	-0.39	0.06	.
Teacher 64	3	25	-0.41	0.29	-0.70	0.17	-0.78	0.04	-0.82	0.26
Teacher 67	1	25	-0.92	-0.05	-0.87	.	-0.45	-0.15	-0.30	.
Teacher 68	5	22	-0.80	-0.41	-0.39	0.40	-0.42	-0.39	-0.03	0.94
Teacher 69	1	24	-0.69	0.00	-0.69	.	-1.04	0.05	-1.09	.

Table A.5: Rubric for Teaching Practice Assessment

Aspect	Score 4	Score 3	Score 2	Score 1
Opening activities	Motivates and provokes curiosity about the upcoming learning activities	Presents learning objectives and motivates students to be interested in the upcoming learning activities	Presents learning objectives without motivating the students to be interested in the upcoming learning activities	There are no opening activities
Presenting learning materials	Presents learning materials in a way that is structured and easy to understand using corresponding examples	Presents learning materials in a way that is structured, clear, and easy to understand, without using examples	Presents learning materials in a structured way that is difficult to understand	Presents learning materials without clear structure and in a way that is difficult to understand
Interaction with students	Conducts two-way communication with students and allows time for student-to-student interaction	Conducts two-way communication with students	Conducts one-way communication with students	Does not make eye contact with students
Teaching aids	Uses teaching aids effectively in helping students to learn, in accordance with the presented materials	Uses teaching aids in accordance with the presented materials, but in a way that is not effective	Uses teaching aids not in accordance with the presented materials	Does not use teaching aids

Attitude	Calm and confident; clear volume; correct tempo	Calm and confident; correct tempo; good volume, but not assertive	Calm and confident; incorrect tempo; low volume	Appears unconfident

Table A.6: Rubric for Teacher Interview Assessment

Teaching Practice Level	Interview Items		
	Teaching Structure	**Differentiation of Teaching**	**Student Assessment**
Level 1	Teacher focuses on a very basic procedural and drilling approach (for example asking students to pray and sing the national anthem at the beginning of the learning session) to deliver learning materials.	Teacher explains the learning materials uniformly to all students regardless of the variation in students' learning progress.	Teacher uses the assessment results only for grading purposes.
Level 2	Teacher focuses on a traditional teacher-centered strategy and drilling approach to deliver learning materials.	Teacher places more attention on students who are lagging behind in their learning progress by providing them with extra lessons.	Teacher uses the assessment results only for grading purposes, while also observing individuals' learning processes.
Level 3	Teacher attempts to use various teaching strategies and media, but remains teacher-centered.	Teacher adjusts their teaching method to accommodate each student based on the individual's learning progress. Lower-ability students receive additional explanations on the materials that they have not understood, while higher-ability students are asked to attempt more advanced learning materials.	Teacher uses the assessment results to improve their teaching strategy.
Level 4	Teacher applies student-centered learning using differentiated instruction.		