Linking Vocational Schools to Industry: Effects on Teachers in Indonesia*

César Garro-Marín[†]
University of Edinburgh

Masyhur A. Hilmy[‡]
University of New South Wales

March 2025

This paper evaluates a mass training program for in-service vocational schoolteachers in Indonesia. The government rolled out an intensive, field-specific professional development program provided by private sector firms to enhance teachers' vocational skills. We use a randomized evaluation to assess its effects on teachers' knowledge, classroom practices, expectations of students' outcomes, and school quality. We find that this program crowded out existing professional development offerings with no increase in overall training participation. There is little evidence of improvements in teachers' knowledge or measures of school quality, albeit with suggestions of increased use of Information and Communication Technologies in the classroom.

Keywords: teacher training, vocational education
JEL Codes: I21, I25, I29

*We thank our partners at Mitrasdudi Kemendikbud for their support and assistance. This paper benefited from feedback and suggestions from Sam Bazzi, Joshua Goodman, Daniele Paserman, and seminar participants at the Boston University Development Reading Group and the 2024 Australasian Development Economic Workshop at the University of Melbourne. We thank Muhamad Haris Zamaludin Setiadiputra for providing outstanding research assistance. Alifia Firliani, Aulia Larasati, Ma'rifatul Amalia, and Terry Muthahhari provided invaluable support for this research project. This research is funded by the Australian Government through J-PAL Southeast Asia – IRF Fund and the Weiss Fund for Research in Development Economics. This project received approval from Boston University Charles River Campus IRB, protocol 5535X and was registered at the AEA RCT registry (AEARCTR-0014062). The views expressed here are those of the authors, and do not necessarily reflect those of the various individuals and organizations here acknowledged.

[†]School of Economics. Email: cgarrom@ed.ac.uk.

[‡]School of Economics. Email: m.hilmy@unsw.edu.au.

1 Introduction

Building a labor force with adequate technical skills is a pressing challenge for education systems in less developed countries (Atkin et al., 2019). Secondary education and, in particular, vocational high schools play a crucial role in addressing this issue. Globally, vocational high schools serve more than 48 million students across low and middle-income countries (EdStats, 2022). Thus, policy interventions seeking to improve their effectiveness could be valuable tools for boosting the employment outcomes of young people in these countries.

In Indonesia, vocational high schools serve more than five million students and comprise half of the enrollment in secondary. In recent years, the Indonesian Government has implemented active policies to increase enrollment in these institutions (Newhouse and Suryadarma, 2011). Despite the policymakers' focus, graduates from vocational high schools face much higher unemployment rates than graduates from the more traditional academic stream. The Indonesian Government identified vocational education quality as one of the factors contributing to this problem.

In this paper, we evaluate a large policy intervention aimed at improving the quality of vocational education in Indonesia. In 2020, the Indonesian Ministry of Education and Culture (MoEC) rolled out a professional development program for vocational high-school teachers called the Upskilling and Reskilling Training Program (UR). This program allocated thousands of teachers across dozens of trainings specific to their vocational field. UR introduced several innovations relative to training programs studied in the literature: (i) it was intensive, with the average course lasting between 6 and 8 weeks;[1] (ii) the trainings were designed and supplied by firms demanding vocational graduates in the labor market, and (iii) it was done at scale, with 2,700 teachers from more than 1,600 different high schools participating in the program. Through these features, MoEC hoped that teachers and schools would align their teaching with the private sector needs while also enhancing the links between vocational schools and potential employers of their graduates.

We worked closely with MoEC and took advantage of the substantial excess demand for the program to conduct a randomized control trial (RCT). We randomly selected applicants to be invited to six vocational training courses related to Information and Communication Technologies (ICT) and surveyed approximately 400 teachers from this sample after the training concluded. These courses met two key requirements: they were oversubscribed, and MoEC had direct oversight of the participant selection and training implementation. Our randomization successfully balanced teachers' characteristics between the treatment and control groups.

We focused our evaluation on understanding whether the training changed teachers' knowledge, their classroom practices, or their expectations of their students' labor market success.[2] We collected data on these outcomes through an original endline survey developed in consultation

[1] This is substantially longer than the 2.5-week median length of the typical teacher professional development program worldwide (Popova et al., 2022).

[2] We focused our evaluation on teachers because they are the primary targets of the UR program, and the primary impacts of the program should arise on changes in teacher behavior first.

with MoEC. In addition, we use MoEC administrative data to study UR impact on school-level measures of quality and student test scores.

We find that rather than increasing teachers' participation in professional development (PD), UR crowded out attendance to PD programs from other providers. Our intervention successfully encouraged participation in UR, with teachers assigned to treatment being nearly twice as likely to participate in UR trainings as teachers assigned to the control group. Additionally, our intervention increased by 16 p.p. teachers' exposure to the private sector (34% of the mean). However, despite these large differences in exposure to UR, we find no difference in overall participation in PD programs between the treatment and control groups. These results suggests that, in the absence of UR, teachers would have attended alternative trainings. In light of this, our estimates should be interpreted as estimates of UR effectiveness relative to the existing options available in the market.

Our evidence suggests UR introduction did not lead to relative improvements in teachers' knowledge, but there are indications of increased ICT use in the classroom. Our ITT estimates for teachers' vocational knowledge are all insignificant and very close to zero. Our 90% confidence interval rules out effects larger than 0.15 standard deviations. At the same time, depending on the specifications, our estimates for ICT use in the classroom may still be consistent with increases between 4 and 15 p.p. in the likelihood of using these technologies to conduct classroom activities.

Analysis of teachers' expectations about their students' outcomes shows that UR made teachers' more optimistic about their students' readiness for the labor market, without corresponding updates on the expected likelihood of employment or their wages. Teachers in the treatment group were 6 p.p. (60%) likelier to rate their students as "industry-ready" than the control group. Meanwhile, the 90% CI of our ITT estimates are sufficiently narrow to rule out increases above US$2.35 per month in the expectations of students' first salary and increases above 2.5 p.p. (0.12 SD) in the expectations of students' employment rate after graduation. Compared to the expected effects in the range of 0.20-0.40 SD that can be attributed to intensive teacher training programs in Fryer (2017)'s review, these results suggest that UR merely updated teachers' beliefs about the alignment of student skills with private sector demands. Moreover, estimates using MoEC administrative data show that UR participation did not lead to improvements in the school accreditation score, a summary of MoEC's assessment of the high school education quality.

Supplementary analysis of our survey responses revealed three possible reasons for the muted impact. First, the training program was not tailored to address teachers' skill gaps, potentially leading to a mismatch between the training implementation and teachers' needs. MoEC's perception of teachers' skills led to the UR's creation, but the training did not target teachers with identifiable weak vocational skills. This led to 80% of attendees reporting that they were already familiar with the materials covered during training. Second, we document that alternative professional development programs remain accessible to teachers in the comparison group, many of which are also vocation-specific. Teachers assigned to the control group reported participating in these alternative trainings, leading to a lack of systematic difference in outcomes with teachers in the treatment group. Finally, after the conclusion of the program, treated teachers reported lacking the support needed to make a sustained change. In particular, only 26% of teachers reported any follow-up sessions after the training and more than half mentioned the need for support from the school principal. This evaluation suggests that in an environment where training opportunities are not scarce, policymakers can instead provide clear guidelines for precisely targeted in-service vocational training that allows sustained collaboration with the private sector instead of launching an upskilling program from scratch.

Our study contributes to the literature of teacher professional development in developing countries. In-service professional development programs for teachers are widespread, but rigorous evaluations of such programs remain scarce. The evidence base on teacher training primarily comes from studies in high-income countries (Fryer, 2017; Yoon et al., 2007). Moreover, evaluations of teacher training programs implemented in low- and middle-income countries have shown mixed effectiveness (Popova et al., 2022). While programs in Argentina and South Africa showed positive effects (Albornoz et al., 2020; Cilliers et al., 2020), different programs in China, Nepal, and Rwanda did not show any effect (Blimpo and Pugatch, 2021; Loyalka et al., 2019; Schaffner, Glewwe and Sharma, 2024), and a teacher training program in Costa Rica led to *worse* student outcomes (Berlinski and Busso, 2017).[3]

We add three distinct contributions to existing studies. First, we provide the first evidence of a program that wanted to improve teaching quality in vocational schools. Policymakers in developing countries place a high priority on vocational education, in marked contrast to international donors who recently redirected their focus toward improving a narrower measure of foundational learning outcomes (Crawfurd et al., 2021). This divergence suggests that efforts to bolster evidence-informed policymaking ought to take into account policymakers' priorities, which calls for more evidence on vocational education. Researchers have run RCTs of vocational programs, but most of these evaluated the effectiveness of short vocational training targeting low-skilled youth (Alfonsi et al., 2020; Attanasio, Kugler and Meghir, 2011; Attanasio et al., 2017) with a few RCTs examining secondary vocational education programs (Field et al., 2019; Hicks et al., 2011). These studies take the quality of vocational education as given, making our study as —to the best of our knowledge— among the first to evaluate interventions aiming at improving the effectiveness of vocational education itself.

Second, we provide evidence from a teacher professional development program that targets the upper secondary level. Until recently, the evidence on teacher training programs in developing countries consisted of interventions targeting primary school teachers (Null et al., 2017).[4] Nevertheless, recent papers have started to add evidence on training for lower secondary teachers:

[3]Other studies bundled teacher training with inputs for students: tablets in Pakistan or textbooks in Papua (Beg et al., 2022; Zaw et al., 2021). A related literature has also investigated the impact of coaching as a form of teacher professional development. Studies include Cilliers et al. (2020) in South Africa, Majerowicz and Montero (2018) in Peru, Yoshikawa et al. (2015) in Chile, and Carneiro et al. (2022) in Ecuador. The first two studies show positive effects on student achievements, while the latter two do not.

[4]Ganimian and Murnane (2016)'s review paper identified only three papers on increasing teacher's skills. Two of them are Abeberese, Kumler and Linden (2014)'s training for fourth-grade teachers in the Philippines and Yoshikawa et al. (2015)'s training for pre-K and kindergarten teachers in Chile.

Berlinski and Busso (2017), Loyalka et al. (2019), and Schaffner, Glewwe and Sharma (2024) evaluated the effect of training programs that target junior secondary math teachers. Blimpo and Pugatch (2021) is a notable exception as they report the results of a comprehensive training program for upper secondary teachers in Rwanda. As developing countries continue to see the increase in (upper) secondary enrollment stemming from the near-universal access to primary schools, building the evidence base on post-primary education becomes a vital priority (Banerjee et al., 2013).

Third, our analysis is based on an evaluation of an intensive (260 hours), subject-specific, and at-scale teacher professional development program. Teacher training associated with specific methods has been highlighted in multiple systematic reviews to improve student learning in developing countries (Evans and Popova, 2016). Subject specificity and multiple-day training are both features of PD programs that are deemed promising to boost student learning outcomes (Popova et al., 2022). At the same time, implementations of promising interventions tested in smaller trials often pose challenges when they are being scaled up or delivered more cheaply. Ganimian (2020) finds null effects for an at-scale intervention on growth mindset, and Kerwin and Thornton (2021) find a weaker effect when a mother-tongue literacy program is delivered at a lower cost. Angrist and Meager (2023) find that variations in the literature of targeted instruction can be attributed to the degree of implementation and program delivery model. Al-Ubaydli, List and Suskind (2019) provide a framework to understand the threats to scaling experiments. With UR's role as an umbrella program to train Indonesian vocational school teachers, our data provides a unique window to look into how teacher training is implemented in diverse vocational streams.

2 Context: Vocational Secondary Schools and the Upskilling-Reskilling training

Indonesian vocational high schools (SMK by its acronym in Indonesian) prepare students for entry into the labor market upon graduation (Pritadrajati, 2018). They service approximately five million students every year and account for about half of the total upper-secondary enrolment in the country. Vocational high schools compete with the General (SMA) and Islamic high schools (MA) to provide students with upper-secondary level education (grade 10-12).[5] The vocational school curriculum places significant emphasis on vocational training, progressively allocating more scheduled time to vocational subjects and internships from grade 10 (26 percent) to grade 12 (72 percent).

Vocational schools offer programs in fields as diverse as performing arts, business, IT, energy, and engineering. The five most popular programs are computer and network technicians, accounting, light vehicle technicians, office administration, and motorcycle technicians. Three-quarters of all SMK in the country offer at least one of these five programs. While some vocational

[5] General schools provide a secular education. Islamic high schools use methods similar to secular schools but teach more religious content (Bazzi, Hilmy and Marx, 2020).

programs are widely available, others are more niche. Programs such as airplane frame construction, Javanese shadow puppetry (*wayang*), crustacean aquaculture, thread manufacturing, and fiberglass boat construction are offered only at handful of schools across the country. Overall, MoEC records 256 unique vocational streams in 14 different fields offered across 14,178 vocational schools (Ditjen Vokasi, 2021).

The Indonesian Government's education policy places a lot of emphasis on vocational education. The number of vocational schools has more than doubled since 2005, and the Government intends to continue with this expansion in the near future (Pritadrajati, 2018). However, despite these investments, recent graduates still face high rates of unemployment. Vocational school graduates aged 30 or less face an unemployment rate that is 45% higher than the average for Indonesians under 30 –19% versus 13%– (Central Bureau of Statistics, Indonesia, 2016). A survey in 2021 by MoEC showed that 48 percent of graduates earn salary below the province minimum wage (Setditjen Vokasi, 2020).

To enhance the competitiveness of its graduates, the Indonesian government initiated a reform of vocational schools aimed at aligning them more closely with industry requirements. In a 2016 presidential decree, the Education Minister was ordered to "link and match" the vocational curriculum with industry needs through partnerships with the private sector (Indonesian Government, 2016). In 2020, MoEC introduced the Upskilling and Reskilling training program to fulfill this objective.

2.1 The Upskilling and Reskilling Training Program

The Upskilling and Reskilling Training Program (UR) was launched by MoEC in 2020. This was a professional development program for vocational school teachers, who taught at schools specialized in five broad industries: hospitality, construction, repair of vehicles and machinery, healthcare services, and software and design services. UR had two primary objectives: (i) improving teachers' vocational knowledge, and (ii) creating links between schools and private sector firms.

UR was comprised of a collection of courses that offered content relevant to the program's target industries. Appendix Table A1 provides illustrative examples of UR course offerings. These courses cover a range of specialized subjects, including photography, 3D animation, and programming and operation of computer numerical control machines (CNC), among others. While each teacher had the opportunity to apply for multiple trainings, they were only able to attend one if selected. The application process was free and conducted online, resulting in minimal application costs.

UR introduced several innovations relative to professional development programs evaluated elsewhere in the literature. First, the program had substantial private-sector involvement because MoEC partnered with firms operating in the target sectors to develop and deliver these courses.[6]

[6] In a review of existing high-quality evaluations of teacher training intervention in developing countries, none

This heavy cooperation with the private sector had two primary objectives: creating links between vocational schools and potential employers for their students and familiarizing teachers with the skills required by private-sector firms. With this in mind, MoEC only set general guidelines for the course format, but the industry partner had ample autonomy to determine the contents. According to the ministry guidelines, potential training providers only needed to meet three broad criteria: (i) being able to organize the training and provide instructors, (ii) having a curriculum and training materials available, and (iii) being able to issue certifications to the training participants (BBPPMPV BOE, 2020).

Moreover, UR provided intensive subject-specific training at scale in more than 50 different vocational subjects. The average training lasted for just over 6 weeks, and, overall, the program trained 2,701 teachers from more than 1,600 vocational high schools spread over 78% (403) of the Indonesian regencies.[7] This contrasts with the typically small professional development program studied in the literature, which usually lasts no more than a couple of weeks (Popova et al., 2022).

UR courses combined online and in-person sessions,[8] with a duration of between six and eight weeks for a total of approximately 260 contact hours. Each course followed a three-phase structure. Firstly, an online phase introduced trainees to the training materials through Zoom and pre-recorded videos. This phase lasted for approximately 30% of the course. Next, the trainees transitioned to in-person class sessions that delivered the contents using the traditional teaching approach. This second phase encompassed approximately 45% of the course. Lastly, teachers proceeded to on-the-job training sessions, where they interned at the facilities of the industry partner. These internships provided teachers with hands-on experience in the day-to-day operations of the company. Following the completion of the training, industry partners awarded certificates to the trainees upon successfully passing an examination. As reference, Appendix Table A2 shows two illustrative course schedules.

To illustrate the type of training UR offered, let us consider the course "*Network and Communications*" where teachers learned basic skills for setting up and maintaining internet networks, e.g., configuring and troubleshooting routers, and configuring data traffic priority for network users (see Appendix Table A3 for a breakdown of the training curriculum). This course was provided by a certified training partner of MikroTik.[9] Upon completion, MikroTik issued certifications to teachers who scored higher than a passing grade (60/100) on the final exam.

Teacher participation in UR was voluntary, but it was greatly encouraged by the government. Besides advertising it on social media, MoEC also sent official invitations to apply to vocational

of the trainings for post-primary education were implemented by private sector firms (Schaffner, Glewwe and Sharma, 2024). Popova et al. (2022) surveyed 33 teacher professional development programs but did not report that any of the programs are designed or implemented by private firms.

[7]Regencies are Indonesian administrative units similar to municipalities in other countries.

[8]This was a necessary adaptation to the Covid-19 mobility restrictions in Indonesia.

[9]MikroTik manufactures a popular brand of network routers and other network equipment. The training for UR participants was provided by PT AsiaVer, its partner based in East Java. Appendix Figures B1 and B2 provide snapshots of the training content and the training activities. MikroTik is a competitor to Cisco, whom MoEC also partnered with to implement other UR training courses.

school principals. Financial costs for the teachers were small: they could apply at no cost. If selected, the entire cost of the training program was borne by MoEC, including transportation, room and board, stipend, and Covid-19 swab tests during the in-person training. Nevertheless, teachers had to obtain an internet plan –if they lacked one– to ensure internet connectivity during the online portion of the training. On average, MoEC incurred on average cost of $2,907 per participant for a total cost of US$7 million in 2020 (Ditjen Vokasi, 2021).[10]

Although participation was voluntary, teachers also had multiple incentives to apply. First, teachers obtained a certification upon successfully completing the training. Second, schools as a whole also had an incentive to be in UR to unlock additional funding from MoEC. The ministry widely publicized that the number of certified teachers in a school would determine the school's eligibility to receive facility upgrading grants.

The selection of the UR participants was largely decentralized to the training providers. In most cases, MoEC gave providers ample discretion with regard to participant selection and the training curriculum. However, MoEC had direct control over participant selection for a subset of subjects which we were able to use for the research design.

Figure 1 illustrates UR's timeline. The teachers' application portal was released in June of 2020, after a delay of several months due to the effects of the Covid-19 emergency in Indonesia.[11] Starting in July, teachers had a two-month window to submit their applications. Then, the trainings took place between October and December of the same year. We collected data through a phone survey at the end of 2021, one year after the trainings concluded.

3 Research Design

Randomized Evaluation. Our main results come from a randomized control trial (RCT) we designed in collaboration with MoEC. This randomized evaluation takes advantage of the substantial oversubscription of the UR training program, with more than 24 thousand applications for only 2,468 available slots. Although the excess demand varied by vocational subject, for most courses there were at least twice as many applicants as slots available (see Appendix Table A4). We were able to randomly select the applicants to invite to the training for a subset of the oversubscribed vocational subjects.

Despite most courses being oversubscribed, MoEC's control over the training implementation varied substantially by subject. For many courses, MoEC delegated the lion's share of the course organization to the training provider, including applicant selection, and syllabus design. This decentralization made it difficult to encourage providers to implement randomized selection of applicants for many subjects.

[10]The budgeted cost was 50% higher than the actual cost. The ministry budgeted IDR137.5 billion (USD 8.9 million) to train 2,160 teachers. Their end-of-year report stated that the actual cost of the program was IDR102.3 billion (USD 6.6 million) to train 2,426 teachers (Ditjen Vokasi, 2021), page 56. Currency conversion uses an exchange rate of IDR 14,500/USD.

[11]Originally, the application portal was intended for release in April of 2020.

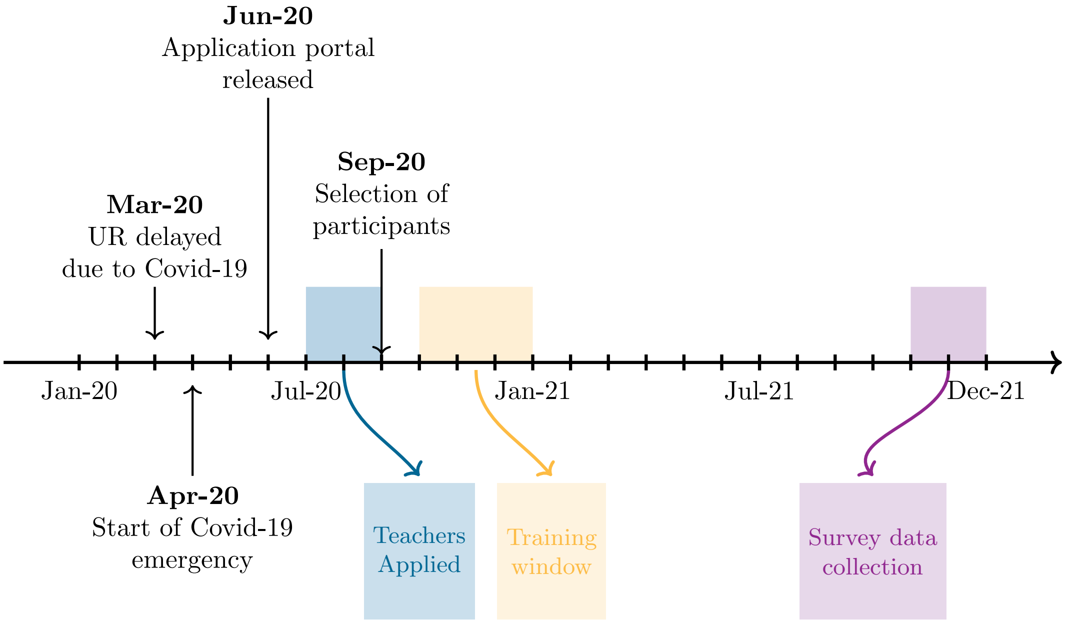


Figure 1: Timeline of the upskilling and reskilling program

Therefore, for our study, we focus on the six subjects where MoEC retained discretion on trainee selection: topography mapping, network and telecommunications, internet of things, 2D animation, Java programming, and database management. Panel A in Appendix Table A4 summarizes the total demand and supply for training spaces for these subjects. For all subjects, training demand substantially exceeded supply. We randomly assigned applicants to be invited to training to match the number of budgeted slots as initially defined by MoEC. The biggest courses were Java programming and Database management, with hundreds of budgeted seats each. For some programs, actual attendees exceeded the budgeted seats because MoEC eventually expanded the number of seats offered.

Our RCT survey sample is comprised of 400 teachers who applied to the UR training, were assigned either to the treatment or control group, and responded to our survey invitation. We collected data from these teachers using a phone survey that we describe in Section 4.1.1. This sample is a subset of the overall randomization frame, where we randomized 844 teachers from 634 schools into treatment and control groups. Of these teachers, 400 teachers responded to our survey.[12] We use the full school sample frame where relevant in analysis with administrative data at the school level.

Matching. We supplemented the RCT design with a sample of nearly 1,100 teachers selected using a propensity score matching (PSM) design. We adopted this alternative strategy for two reasons: first, to achieve greater precision with a larger sample size, and second, to compare its results with those of the RCT. Similar results under both strategies would provide additional support to the results from the RCT. Panel B in Appendix Table A4 lists the 15 courses in the PSM design with the most applicants. We selected the PSM sample by matching treated

[12] Respondent teachers and non-respondent teachers have overall similar characteristics; however, non-respondents were more likely to have numbers associated with smaller telecommunication carriers.

teachers to suitable control applicants using pre-treatment individual-level data from MoEC administrative databases. We used information about applicants' gender, education level, province of residence, teaching experience, and competency test score, among others. Further details on the matching approach are given in Appendix C.

4 Data

4.1 Data sources

We use two main data sources: an original phone survey that we use for our main results, and administrative data from MoEC.

4.1.1 Original survey data

We collected data on post-treatment outcomes through a survey deployed during November and December of 2021. We conducted the survey via phone to minimize the challenges posed by the spread of Covid-19 and the logistical costs of reaching teachers spread across the whole Indonesian archipelago.[13] We collected data on three main outcomes: teachers' vocational field knowledge, their classroom practices, and their expectations about their students' labor market success. Besides these outcomes, we also collected detailed data on teachers' professional trainings, as well as basic demographic information.

Teachers' vocational knowledge and classroom outcomes are our primary outcomes. Any other impacts on, for instance, students' labor market outcomes would arise only as a result of improvements in teachers' quality or their teaching practices. Alternatively, they could also arise from the second-order effects of strengthened connections between schools and employers.

We measure improvements in teachers' vocational knowledge with a battery of training-specific true-false questions that test the contents taught in the UR trainings. We developed ten questions for each of the six courses in the RCT design based on the training materials and the post-course test that the training partners administered to the participants. We modeled these questions after the multiple-choice tests employed by Alfonsi et al. (2020) in their vocational training experiment in Uganda,[14] but simplified to fit the constraints imposed by the phone survey format. We validated them with MoEC staff in charge of the training implementation.

[13] The spread of our respondents in 384 districts over 36 provinces (i.e., 75% of all districts and 95% of all provinces) would make an in-person survey prohibitively expensive.

[14] In particular, we referred to their sample of multiple choice questions to measure sector skills test (in their Fig A5, the skill test for motor mechanics). More broadly on the use of phone assessment to measure knowledge, Angrist et al. (2020) argued that oral assessment of learning are readily adaptable to phone surveys as part of an RCT in Botswana. To test the reliability and validity of their measures, they found a correlation of 69% between simple oral questions and a comprehensive assessment from ASER. The correlation represents a reasonable concurrent validity, which builds overall construct validity of phone-based assessments. See also the use of phone surveys to measure learning in Kenya by Rodriguez-Segura and Schueler (2022), who demonstrated that phone-based assessment performed well as an aggregate measurement in the context of impact evaluations.

The validation was done in several stages, and we tested the survey module extensively prior to the survey's launch to ensure that respondents could answer the questions over the phone.

In Table 1, we show that our knowledge questions capture meaningful variation in teachers' skills. The table correlates our test scores with the standardized scores in the 2015 Teacher Competency Exam (UKG) from the administrative data. UKG scores are used by MoEC as a measure of teacher ability (see Section 4.1.2). Table 1 shows that scores in our test are positively correlated with both the overall UKG scores, as well as with the vocational and pedagogical components separately. In addition, Appendix Table A5 summarizes teachers' responses to our knowledge test by question. This table shows the questions were neither too hard nor too easy for the participants, as only 9 of the 60 questions were answered correctly by more than 90% of the teachers. Moreover, very few respondents gave a uniform answer to all statements.

Our measures of teaching practices come from a survey module capturing teachers' time use in the classrooms, the equipments they used to teach, as well as their teaching load. Previous research links teachers' time distribution across activities to student achievement (Jukes, Vagh and Kim, 2013; Stallings, 1980). UR placed strong emphasis on hands-on learning, which could have encouraged teachers to modify their teaching. Our instrument is informed by the Stallings classroom observation instrument, a tool widely used in developed nations to assess teachers' effectiveness in managing their classrooms (Bruns, De Gregorio and Taut, 2016; World Bank, 2015). They are also similar to the instructional time questions in the National Teacher and Principal Survey (NTPS) from the US National Center for Education Statistics (NCES, 2021).

We proxy students' downstream outcomes with information about teachers' expectations of their students' labor market outcomes. Specifically, we ask them to estimate the share of their students who are employed three months post-graduation, their average salary, and their university enrollment rates. We ask them to estimate these outcomes for students graduating in May 2021 (after the treated teachers participated in the UR training) and May 2019 (prior to the program and also prior to the pandemic, which caused severe labor market disruption). While teachers' beliefs are a noisy measure of students' outcomes, in Appendix Table A6, we show that they capture meaningful variation in students' outcomes. The table regresses multiple measures of actual student outcomes on teachers' expectations. Teachers' answers for the 2019 graduates are strongly correlated with that year's national exam scores in math, Indonesian, English, their vocational subjects, and the overall average score. Given that the national exam was cancelled during the Covid-19 pandemic, teachers' expectations for the 2021 graduates provided us a practical proxy of student outcomes.

We also collected detailed data on professional training activities using questions modeled after the In-Service Teacher Training Survey Instrument (ITTSI) questionnaire in Popova et al. (2022) to systematically capture implementation characteristics. These characteristics allow us to investigate how training organization, content, and delivery may influence the outcomes.

Finally, we note that we do not expect residual biases (e.g., social desirability bias) to differ systematically between the treatment and comparison group in the survey. The administration of the survey by an independent company not associated with MoEC helps to minimize this concern among respondents. An established company with experience in phone surveys contacted the teachers on behalf of the research team using the teachers' phone number in the MoEC application database. Research staff at J-PAL SEA ran monitoring and quality assurance steps during the data collection period to ensure the fidelity of the recorded answers (backchecks, high-frequency checks, and spotchecks).

4.1.2 MoEC administrative data

Our main administrative data comes from three databases: the 2015 Teacher Competency Exam (UKG by its acronym in Indonesian), the School Accreditation and the National University Entrance Exam databases. We use the UKG database to obtain pre-treatment individual-level data about teachers, which allows us to describe the kind of people UR attracted. Additionally, the entrance exam and the school accreditation datasets provide us with school-level measures of quality before and after the UR training.

The 2015 UKG database contains individual-level information covering the universe of active school teachers in 2015. It contains data on demographic characteristics (age, education, gender, place of residence, etc), employment characteristics (school of employment, vocational specialty), and the teacher's scores in the exam.

The UKG Exam was initially administered as a part of the teacher certification process that allows teachers to unlock a salary supplement. However, in 2015, MoEC ran a nationwide exam to measure and identify gaps in teachers' quality across the country (Menteri Pendidikan dan Kebudayaan, 2015). Teachers were tested with a two-hour exam on two areas: proficiency in their teaching subjects (70%) and their pedagogy skills (30%). Teachers scoring below the passing threshold (55%) were referred to a remedial professional development program.

We successfully matched 67% of applicants (66% of applications) to their test score data. Appendix Table A7 describes the UR applicants according to their match status with the test score data. Matched applicants are older, and they are slightly more likely to live in Java. Because the UKG data only contains teachers who were active in 2015, it makes sense that we lack information from younger teachers. For our survey respondents, Appendix Figure B3 shows the distribution of their UKG scores in our data.

We obtain school-level outcomes from both the School Accreditation database and the College Entrance Exam rankings, which provide us with direct observation of school-level outcomes not dependent on teachers' beliefs. We obtain measures of school quality from the School Accreditation Archive (BAN-PDM, 2025). School accreditation is a process through which MoEC's National Accreditation Board (BAN-PDM by its Indonesian acronym) evaluates school quality and classifies schools into four categories: A (highest), B, C, and not accredited. The rating is assigned based on student outcomes, teacher quality, and school management quality. We collected the accreditation ratings for the schools in our sample before and after 2020 (UR training year). As an alternative measure of quality, we also use data from the schools' rank in the National College Entrance Exam (Higher Education Entrance Test Institute, 2022). This is a high-stakes exam that all students wishing to continue to university must sit. The test administration agency publishes the names of the top 1000 schools in the country, along with the average school score. We match schools by name to the 2020 and 2021 rankings and use a dummy of being in the top 1000 as a dependent variable.

We also use individual-level data from the National Assessment (AN by its Indonesian Acronymn) as an additional measure of student outcomes (MoEC, 2023). The National Assessment is a low-stake evaluation introduced in 2021 after the discontinuation of the high-stake national examination that students had to take as a requirement for graduating high school. Before 2021, grade 12 SMK students took exit exams in four different subjects: math, Indonesian, English, and vocational subjects. In the National Assessment, grade 11 students take low-stake exams that assess their literacy and numeracy skills. We matched the schools in our evaluation sample with the schools in the AN public use microdata sample (PUMS) using exact matches on the district name and school status, and fuzzy matches on the number of classrooms and number of students. This procedure left us with approximately 1,200 students in our sample.

4.2 Summary statistics

Table 2 summarizes the characteristics of the sample we use in our RCT design. Column (1) shows statistics for the whole sample, while Columns (2) and (3) present means by treatment status assignment. In addition, Column (4) shows the difference in means between the treatment and the control groups after netting out vocational sector (strata) fixed effects.

The typical teacher in our sample has a bachelor's degree, is employed full-time, has taught for approximately 10 years and earns US$242 per month. They teach classes in multiple grades, primarily in the Programming and ICT subjects (81%) and entrepreneurship (14%). Our sample is spread across 32 provinces and has good geographical representation: 52% and 24% are located in Java and Sumatra, which roughly corresponds to these islands' share in the Indonesian population. Overall, the treatment and control groups are well-balanced across nearly all characteristics. With respect to gender, treated teachers are 8 p.p. less likely to be men, but this difference disappears once we account for randomization strata.[15]

Appendix Table A9 summarizes the characteristics of all the UR trainings included in our survey, as reported by respondents. The average UR training lasted for 6.8 weeks (274 hours) and about 90% of trainees had contact with a trainer from the private sector. Other than activities such as lectures and discussions, participants also reported direct skill-building activities such as practice sessions with computers (60%), internship at the industry partner (20%), and a pedagogical component in the form of teaching practices (28%). Teachers also reported receiving materials

[15] In Table A8 we show the characteristics of UR attendees and their matched controls for the PSM design. Overall, the matching was successful at balancing the characteristics across both groups. Other than the total number of classes taught and the share living in the outer islands, all of the differences between treated teachers and their matched controls are small and insignificant.

from the training (83%) and nearly half received lesson plans/videos. The most cited benefits from the trainings were obtaining a certification from the industry partner (66%) and increases in knowledge and skills (59%). The top two training components rated as most helpful were on-the-job training (29%) and the training material (24%). After the conclusion of the training, 87% of teachers reported they incorporated training content in their day-to-day teaching and 78% reported sharing training materials with other teachers in their school. Three in five participants reported having proficiency in some of the training materials prior to the training, while one in five reported knowing all materials prior to the training, suggesting there is room for improvements in both the selection process and the training syllabus. Overall, teachers report high rate of satisfaction with an 8.8 average score on a scale of 10.

4.2.1 Who applied to UR?

Upon the launch of the UR program, the Ministry invited schools and teachers to apply for the training. The eligibility criteria advertised in the MoEC-issued guidelines, official letters, and YouTube live-stream launch were by no means restrictive. Schools needed to have at least two people teaching in the vocational sector, and they needed to be able to guarantee that students' learning would continue while the selected teachers participated in the UR training. The requirements for teachers were similarly broad. Participating teachers needed to be registered in the MoEC database, have at least a college degree, be no older than 50, be teaching a vocational subject in their schools, and be willing to apply the training materials in their original schools upon completion. The first two of these are basic requirements that any teacher must meet to be allowed to teach at a school.

UR was effective at attracting teachers and schools. In total, 32% of all vocational schools in the country had at least one teacher submitting applications. Teachers could apply to several training subjects, and the average participating school submitted 2.55 applications from 2.42 teachers. These teachers represent 16.5% of all the teachers teaching vocational subjects nationwide.

The availability of administrative data for the universe of teachers allows us to describe the kind of teachers UR attracted. In Table 3 we matched the MoEC's application roster to the 2015 UKG database and restricted the sample to teachers aged 23 to 45 in 2015. Then we regress a dummy equal to one if the teacher is a UR applicant on a series of individual and school characteristics, and the standardized test scores. All regressions include province and vocational field fixed effects and column (3) additionally controls for teachers' alma mater. Column (3) in Table 3 shows that UR attracted younger and less experienced teachers, working in public schools and with permanent contracts. Male teachers applied at higher rates, with women being 1.7 p.p. (20% of the mean) less likely to apply.

UR applicants were positively selected on ability. Table 3 controls for the standardized scores in the pedagogical and vocational components of the 2015 Competency Exam. UR applicants performed better in both areas, with stronger positive selection on vocational aptitude. The coefficient in column (3) indicates that scoring one standard deviation higher in the vocational test is associated with an increase of 1.5 p.p. in the probability of applying (19% relative to the mean). In comparison, although still positive, the pedagogy score coefficient is less than half that size.

5 Upskilling and Reskilling evaluation

We evaluate the effects of the UR on teachers' outcomes and expectations by comparing the results of individuals assigned to treatment and control as follows:

$$Y_i = \alpha + \beta Treated_i + X_i \gamma + \delta_f + \varepsilon_i \quad (1)$$

where Y_i denotes the outcome of interest, $Treated_i$ is an assigned-to-treatment dummy, and X_i denotes additional controls that might be included. Because we stratified the randomization, we include vocational field fixed effects δ_f in the specification.

In anticipation of the analysis, we registered a pre-analysis plan at the launch of the phone survey, prior to the completion of the survey data. Our pre-analysis plan is registered at the 3ie's Registry for International Development Impact Evaluations (RIDIE) platform, which allows the registration of studies using randomized evaluation and quasi-experimental designs.[16] The primary outcomes were teachers' vocational knowledge, classroom outcomes, and teachers' expectations of their students' labor market outcomes, as described in Section 4.1.1. In addition, we present results that use school-level outcomes from MoEC's administrative data.

Changes in budget allocations and several other implementation considerations by the MoEC led to adjustments in the number of teachers invited into the UR training (See Appendix Table A4). These changes influenced the training attendance among our sample, leading to imperfect compliance. Therefore, for our main results we report the estimates from intention-to-treat analysis (ITT).

In addition to the ITT analysis for the six sectors, we also report the following estimates: (a) ITT excluding the worst-compliance vocational training, (b) 2SLS, (c) 2SLS excluding the worst-compliance vocational training, (d) Per-protocol subpopulation, and (e) Propensity score matching (PSM). For the 2SLS estimates, we use the initial assignment as an instrumental variable for attendance in the course. We report the ITT and 2SLS estimates excluding the worst-compliance vocational training to improve precision with the existing sample. Our additional exploratory analysis includes estimates using the per protocol (PP) subpopulation, i.e., the set of respondents who adhered fully to the assignment status. The PP analysis can be conceptualized as estimating an answer to "what is the effect of actually receiving a treatment if adherence to study protocol could be improved?"—which is different to "what is the effect of assigning a treatment?" that the ITT analysis answers.[17] The PSM analysis expands the

[16] Study ID: RIDIE-STUDY-ID-619b30d3ad31d, accessible at https://ridie.3ieimpact.org/

[17] In the clinical trial literature, Tripepi et al. (2020) highlighted the guidelines from the European Medicine

sample size and adds teachers in more vocational sectors.

5.1 UR's impact

We start by showing that the intervention successfully increased teachers' participation in UR trainings, but surprisingly, with no increase in overall participation in professional development courses. In Table 4, we report effect estimates for several measures of teachers' training activities. Panel A, columns (1) to (3) show that assignment to treatment increased UR training participation and exposure to private firms. Being assigned to treatment increased the likelihood of program participation by 21 p.p. and the likelihood of receiving training by a private firm by 17 p.p. However, the intervention did not increase overall training participation or training hours. Panel A column (4) and Panel B columns (1) to (4) show there are no significant differences between treated and control teachers in attendance to –any– training or the number of hours spent in training. This suggests that rather than making teachers more likely to engage in professional development activities, UR likely shifted their attendance away from other existing teacher professional development programs.

We now study whether UR had meaningful effects on teachers' vocational knowledge, classroom practices and expectations of student outcomes. Because our main results rely on an endline survey, they compare post-treatment outcomes between teachers assigned to the treatment and control groups. As teachers in the control group also engaged –rather enthusiastically– in professional trainings, our ITT estimates answer the question of whether UR generated an improvement in these outcomes relative to a teacher engaging in the typical professional development activities. Although these estimates do not directly answer whether UR was effective at increasing teachers' knowledge –relative to a no-training scenario– they are still informative on whether the additional expense incurred with UR represented an improvement over the existing training offerings.

Our analysis focuses on four groups of outcomes: (i) vocational knowledge, (ii) classroom practices, (iii) teachers' expectations of student outcomes, and (iv) alternative measures of school quality. The first three use data from our original survey, while the latter uses MoEC administrative data. Overall our results suggest that UR-exposed teachers did not make meaningful relative improvements in their vocational knowledge. Yet, there is some indication that they use more Information and Communication Technologies (ICT) in the classroom. UR-exposed teachers also became more optimistic about the readiness of their students for the labor market, with no change in their expected salary and employment rates. Moreover, we find no evidence of improvements in school outcomes coming from administrative data.

In Table 5, we show ITT estimates for vocational knowledge and classroom practices. The UR trainings included in our sample focused on technical content in ICT-heavy sectors and

Agency's Committee for Proprietary Medicinal Products, which stated that both ITT and PP results should lead to similar conclusions for a robust interpretation. See also Ye et al. (2014) for a simulation study from the medical literature and Peugh et al. (2017) from the psychology literature discussing the use of PP analysis vis-a-vis ITT in estimating treatment effects in RCTs with non compliance.

emphasized hands-on learning. If UR were more effective than the existing offerings it replaces, we would expect improvements in vocational knowledge and heavier use of ICT technologies within the class (Ditjen Vokasi, 2021). Nevertheless, the estimated coefficient for vocational knowledge in Column (1) is small and very close to zero.

Because of the less-than-perfect compliance with treatment assignment, we also report per-protocol effect estimates in panel B of Table A10. Column (1) shows a significant per-protocol estimate of 3.6 p.p. (0.4 questions or 0.27 S.D.), which could be consistent with some knowledge gains by the attendees. Nevertheless, per-protocol estimates are informative of treatment effects in a better-compliance scenario only under the *very strong* assumption that compliance is uncorrelated with potential outcomes. In our context, compliance is unlikely to be fully exogenous. In Appendix Table A11, we relate UR attendance to teachers' characteristics separately for teachers assigned to the treatment and control groups. Actual attendees seem more likely to be permanent employees, live in urban areas and, among the treated, have lower UKG scores. Therefore, we view the per-protocol estimates merely as suggestive.

Columns (2) to (4) in both panels of Table 5 present our main effect estimates for teachers' classroom practices. Panel A shows effects for the share of teaching time spent in various classroom activities, while panel B shows estimates for the share of teachers using ICT to conduct classroom activities. There is little evidence that UR-exposed teachers changed the way they distributed their lesson time relative to control teachers, as all the point estimates are small and insignificant. Nevertheless, panel B could still be consistent with increased ICT use in UR-exposed teachers. Although imprecise, the 5.9 p.p. estimate for ICT use for class discussion is sizable relative to the 36% sample mean. Estimates that exclude the worst-compliance sector in panel A of Appendix Table A12 suggest positive but imprecise increases of approximately 4 p.p. in ICT use to cover material and pupil work. In addition, the per-protocol estimates suggest a significant increase of 15 p.p. in ICT use for discussion and imprecise increases of 4 and 6 p.p. in ICT use to cover material and pupil work respectively.[18]

In Tables 6 and 7 we turn to the study of student outcomes. Table 6 shows ITT estimates for outcomes coming from a battery of questions on teachers' expectations about their student labor market outcomes. Columns (1) to (3) show little changes in students' expected employment rate, wages, and university attendance. The negative estimate on wages rules out positive wages effects over $4.19 (0.06 SD) with 95% confidence. However, column (4) shows that UR-exposed teachers became more optimistic about their students' preparedness for employment in the private sector. Treated teachers were 6.3 p.p. more likely to rate their students as ready for employment in firms in their vocational sector. This suggests that UR led teachers to update their beliefs about how their students' skills aligned with the labor market demands without a corresponding change in their beliefs about their overall success. Notably, these patterns also arise in the 2SLS and per-protocol estimates in Appendix Table A13.[19]

[18] Appendix Table A10 also report 2SLS estimates for classroom practices, but they are too imprecise to draw meaningful inferences.

[19] Because Table A11 shows that UR attendees tend to have lower scores in the vocational portion of the Teachers' Competency Exam (UKG), we also produced estimates that control for the vocational UKG score (not shown). Including the UKG score as a regressor reduces our sample by about 8% (31 respondents) because not

In Table 7 we show results for outcomes coming from MoEC's administrative data. The main limitation of the results in Table 6 is that they use teachers' expectations as main outcomes, which are still a noisy measure of students' actual outcomes despite being strongly correlated with them. Nevertheless, in Table 7, we find no evidence of any effect from UR when using MoEC's administrative data on school quality. In columns (1) to (4), we use data from the MoEC's school accreditation program to study whether UR participation led to improvements in school accreditation scores. MoEC's accreditation program assesses school quality and classifies schools into four categories: A (best), B, C (worst), and nonaccredited. In columns (1) and (2) we code the school's accreditation level continuously –with higher values representing better quality– and regress them on a dummy of whether the school had an assigned-to-treatment teacher. In columns (3) and (4), we perform an analogous exercise using as outcome an indicator of A-level accreditation. All point estimates are negative and insignificant, indicating that UR participation did not lead to updates in MoEC's quality assessment for UR-participating schools. Additionally, in columns (5) and (6), we use data from the high-stakes Indonesian University Entrance Examination and use an indicator of being in the top 1000 schools with the highest average test scores nationwide as the outcome, while in Appendix Table A14 we use student-level test score data from the literacy and numeracy components from the 2021 Indonesian National Examination. We also find null effects on all test-score outcomes, although admittedly, the contents taught in UR were technical, and they likely do little to improve students' preparedness for university or their numeracy skills.

We also examine whether the program had differential effects depending on the teachers' characteristics. In Table 8, we consider two main outcomes: teacher knowledge and graduates' employment expectations. We consider the following teacher characteristics: having a master's degree or higher, having teaching experience above the median (12 years), being a permanent employee, having MoEC certification, being a male teacher, and being from Java/Bali. We find no consistent pattern suggesting that UR was particularly effective for a subgroup of teachers. For the estimates in vocational knowledge in Panel A, the coefficients for treatment indicators and their interaction with the group indicator are fairly small, suggesting that any possible effects will be smaller than 1 question. A similar picture arises in results for teachers' expectations, as shown in Panels B–C.

Finally, our results using a Propensity Score Matching design in a larger sample of vocational sectors produce results similar to those in our RCT. Panel A Appendix Table A15 shows null effects on time spent on various classroom activities coupled with positive effects on ICT use in the classroom in panel B. In addition, column (3) in panel C indicates that participating teachers became more optimistic about their students' labor market readiness. Nevertheless, unlike the RCT sample, columns (1) and (3) suggest that UR-exposed teachers expect higher employment and lower university entrance rates from their students.

all our respondents took the exam in 2015, but there is little change to our point estimates. Therefore, negative selection in teachers' vocational skill does not seem to explain the small UR effect estimates.

6 Discussion on UR's Small Impacts

We have documented that the program did not lead to transformative improvements for the treated teachers vis-a-vis the comparison group. In this section we explore three possible reasons for this finding.

6.1 Mismatch between teachers' needs and training offerings

For a teacher training program to be effective in improving student outcomes, it needs to improve the teachers' knowledge and classroom practices. If teachers are not at all familiar with the curriculum, then training programs with more basic contents or refresher courses targeted at specific knowledge and skill gaps could be a first step to improve teachers' effectiveness. In other words, the theory behind effective Teaching at the Right Level interventions could be extrapolated to the teaching force to address teachers' skill gap.

MoEC is aware of a general skill gap among vocational teachers and sees the high unemployment rate among graduates as a symptom of quality issues in vocational high schools. However, addressing the skill gap with an at-scale program for such a diverse educational system is challenging. Teachers' responses to our survey, along with reports from the training, give clues as to why training mismatch may have contributed to the lack of impact of the program.

MoEC selection guidelines were broad and not targeted. Moreover, none of the UR attendees we surveyed perceived that the selection processes were targeted to address possible skill gaps. 47% reported that their school was selected because of the vocational sector they offered, 25% reported that there were no particular selection criteria, and 20% reported that selection was based on who submitted an application to MoEC's portal. Any targeting within school, if any, was coarse and did not go beyond their teaching duties: 72% of teachers reported being selected for training (or being selected by the school to submit an online application) based on what subjects they taught at school. 28% of teachers reported that there is no within-school selection.

Moreover, teachers were also likely to report that they were already familiar with the materials delivered during the training. Eighty percent of the attendees reported that they had taught the materials to the students prior to the training. They also reported some degree of proficiency prior to the training: more than three-fifths of attendees reported that they were already proficient in some of the contents, and a full one-fifth reported proficiency in all of the contents. Respondents in our sample were also able to identify topics from the training that they had been teaching to their students. For example, multiple attendees in the Java programming training mentioned 'foundational Java', and attendees in the fiber optics training mentioned 'DHCP server' as materials they regularly taught at school. These materials may be the same materials that UR training had covered. These findings could also help explain why we see an increase in teachers' optimism about their students without any meaningful change in their subject knowledge or teaching practices. If teachers received content they were already familiar with from the training providers, they could have interpreted this as evidence that they were already teaching skills demanded in the private sector.

6.2 Outside training provided an alternative in counterfactual

A training program could be successful in increasing the participants' skills if they do not have access to comparable training in the counterfactual. In this context, however, teachers could (and did) access alternative training outside the UR scheme. Descriptive lists of trainings that non-UR participating teachers provide during our phone surveys reveal that various institutions offered trainings to teachers beyond the UR scheme. Among our respondents, 63 respondents report that they received training on utilization of e-learning platforms/Covid distance learning adjustments and 26 respondents listed trainings that are specific to a vocational sector as well. Examples from the latter group include training in Python programming language, IP address rooting, CSS and Javascript for web programming, CAD, welding, and machinery techniques. In this light, it may not be required for MoEC to implement the training on their own, so long as they provide clear guidelines for the private sector to collaborate with vocational high schools in a bid to improve the quality of the vocational education system in the country.

6.3 Lack of post-training support

Centrally organized training brings teachers away from their daily environments where they deliver the curriculum to the students. For teachers to be able to apply the training that they received, they may require further support after the training was concluded. This may include following up with teachers or obtaining authorization from the school principal to incorporate the materials that they received from the training into their day-to-day teaching.

However, providing training follow-up remains a best practice that is rarely implemented. More than half (55%) of the teachers in our sample who attended UR reported that they needed further support to be able to incorporate training into classroom practices. At the same time, only a minority of teachers recalled any follow-up sessions from the training. The overwhelming majority (74%) did not receive any follow-up. The lack of post-training support has been argued as one of the explanations for the lack of impact for an at-scale training program for middle-grade teachers in Nepal (Schaffner, Glewwe and Sharma, 2024). In comparison, Popova et al. (2022) noted that 85% of the top performing programs in their data include follow-up visits, while only 49% of the at-scale programs they analyzed include a follow-up visit.

Furthermore, teachers who found the training useful may also have to navigate negotiations with school principals. Among teachers attending the training, 53% reported that they would have needed management support from their principals to incorporate the materials from the training into their classroom practices. Slightly less than half of the teachers (47%) reported that they took steps to coordinate with their school principals. Without buy-in from the principals, this may have led to the lack of meaningful changes in teachers' practices in the classroom.

Hands-on training with industry may also reveal the infrastructure gap between industry and the teachers' vocational schools. Teachers may gain access to specialized equipment and software during the training, but the same facilities may not be available to the teachers to use in the classroom. Accordingly, 60% of teachers said that they needed specialized equipment to incorporate materials from the training, while also highlighting students' need to access computers (55%), specialized software (41%), and internet access (48%). In a setting where only 70% of students have sufficient internet access, adaptation becomes challenging.

7 Conclusion

As education policymakers in developing countries prioritize vocational education, improving its effectiveness holds great potential to create meaningful impact for their students (Crawfurd et al., 2021). Teacher professional development programs that bring vocational teachers closer to the private sector have strong theoretical appeal, but challenges remain to implement an effective PD program at scale. This study finds that an at-scale intensive teacher PD program for vocational teachers in Indonesia did not have any transformative impacts on teacher knowledge, teaching practices, and expectations of their graduates. Our evaluation adds to the PD literature that finds little impact of at-scale programs when they are being rigorously evaluated (Loyalka et al., 2019; Popova et al., 2022; Schaffner, Glewwe and Sharma, 2024).

Our study makes three major contributions to the PD literature. First, we provide the first rigorous evidence of a program to improve teaching quality in upper secondary vocational schools. Second, we evaluated an at-scale PD program which serves as the umbrella program to train teachers in dozens of diverse vocational streams. Third, our analysis is based on an evaluation of an intensive program (260 hours) and subject-specific training, both of which are features of PD programs that are deemed promising to boost student learning outcomes.

Our evaluation offers valuable lessons from Indonesia to other policymakers interested in designing their teachers' professional development programs. Participating teachers' survey responses highlight the importance of a needs assessment, which may help align interventions to target existing skill gaps better. While our findings are rooted in the specific context in which this program was implemented, our evaluation offers a rare case through which other policymakers wanting to improve their vocational education systems can build upon.

References

Abeberese, Ama Baafra, Todd J Kumler, and Leigh L Linden. 2014. "Improving Reading Skills by Encouraging Children to Read in School:: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines." Journal of Human Resources, 49(3): 611–633.

Albornoz, Facundo, María Victoria Anauati, Melina Furman, Mariana Luzuriaga, Maria Eugenia Podesta, and Inés Taylor. 2020. "Training to teach science: experimental evidence from Argentina." The World Bank Economic Review, 34(2): 393–417.

Alfonsi, Livia, Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali. 2020. "Tackling youth unemployment: Evidence from a labor market experiment in Uganda." Econometrica, 88(6): 2369–2414.

Al-Ubaydli, Omar, John A List, and Dana Suskind. 2019. "The science of using science: Towards an understanding of the threats to scaling experiments." National Bureau of Economic Research.

Angrist, Noam, and Rachael Meager. 2023. "Implementation Matters: Generalizing Treatment Effects in Education." SSRN 4487496.

Angrist, Noam, Peter Bergman, David K Evans, Susannah Hares, Matthew CH Jukes, and Thato Letsomo. 2020. "Practical lessons for phone-based assessments of learning." BMJ Global Health, 5(7): e003030.

Atkin, David, Dave Donaldson, Imran Rasul, Matthieu Teachout, Eric Verhoogen, and Christopher Woodruff. 2019. "Firms, trade, and productivity." International Growth Centre, London.

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." American Economic Journal: Applied Economics, 3(3): 188–220.

Attanasio, Orazio, Arlen GuarÍn, Carlos Medina, and Costas Meghir. 2017. "Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-Up." American Economic Journal: Applied Economics, 9(2): 131–143.

Banerjee, Abhijit, Paul Glewwe, Shawn Powers, and Melanie Wasserman. 2013. Expanding Access and Increasing Student Learning in Post-Primary Education in Developing Countries: A Review of the Evidence. Cambridge, MA:Abdul Latif Jameel Poverty Action Lab.

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx. 2020. "Religion, Education, and the State." National Bureau of Economic Research Working Paper 27073.

BBPPMPV BOE. 2020. "Panduan Pelaksanaan Diklat Program Upskilling dan Reskilling Guru Kejuruan pada SMK Berstandar Industri Tahun 2020." Balai Besar Pengembangan

Penjaminan Mutu Pendidikan Vokasi Bidang Otomotif dan Elektronika Direktorat Jenderal Pendidikan Vokasi Kementerian Pendidikan dan Kebudayaan.

Beg, Sabrin, Waqas Halim, Adrienne M Lucas, and Umar Saif. 2022. "Engaging teachers with technology increased achievement, bypassing teachers did not." American Economic Journal: Economic Policy, 14(2): 61–90.

Berlinski, Samuel, and Matias Busso. 2017. "Challenges in Educational Reform: An Experiment on Active Learning in Mathematics." Economics Letters, 156: 172–175.

Blimpo, Moussa P., and Todd Pugatch. 2021. "Entrepreneurship Education and Teacher Training in Rwanda." Journal of Development Economics, 149: 102583.

Bruns, Barbara, Soledad De Gregorio, and Sandy Taut. 2016. "Measures of Effective Teaching in Developing Countries." RISE Working Paper 16/009.

Carneiro, Pedro, Yyannu Cruz-Aguayo, Ruthy Intriago, Juan Ponce, Norbert Schady, and Sarah Schodt. 2022. "When promising interventions fail: Personalized coaching for teachers in a middle-income country." Journal of Public Economics Plus, 3: 100012.

Central Bureau of Statistics, Indonesia. 2016. "National Labor Force Survey (Sakernas)."

Cilliers, Jacobus, Brahm Fleisch, Cas Prinsloo, and Stephen Taylor. 2020. "How to improve teaching practice?: an experimental comparison of centralized training and in-classroom coaching." Journal of Human Resources, 55(3): 926–962.

Crawfurd, Lee, Susannah Hares, Ana Minardi, and Justin Sandefur. 2021. Understanding Education Policy Preferences: Survey Experiments with Policymakers in 35 Developing Countries. Center for Global Development.

Ditjen Vokasi. 2021. "Laporan Kinerja Direktorat Jenderal Vokasi Tahun 2020." Direktorat Jenderal Pendidikan Vokasi Kementerian Pendidikan dan Kebudayaan.

EdStats. 2022. "Education Statistics (https://datatopics.worldbank.org/education)."

Evans, David K, and Anna Popova. 2016. "What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews." The World Bank Research Observer, 31(2): 242–270.

Field, Erica M., Leigh L. Linden, Ofer Malamud, Daniel Rubenson, and Shing-Yi Wang. 2019. "Does vocational education work? Evidence from a randomized experiment in Mongolia." National Bureau of Economic Research w26092.

Fryer, Roland G. 2017. "The production of human capital in developed countries: Evidence from 196 randomized field experiments." In Handbook of economic field experiments. Vol. 2, 95–322. Elsevier.

Ganimian, Alejandro J. 2020. "Growth-mindset interventions at scale: Experimental evidence from Argentina." Educational Evaluation and Policy Analysis, 42(3): 417–438.

Ganimian, Alejandro J, and Richard J Murnane. 2016. "Improving education in developing countries: Lessons from rigorous impact evaluations." Review of Educational Research, 86(3): 719–755.

Hicks, Joan Hamory, Michael Kremer, Isaac Mbiti, and Edward Miguel. 2011. "Vocational education voucher delivery and labor market returns: A randomized evaluation among Kenyan youth." Report for Spanish Impact Evaluation Fund (SIEF) Phase II.

Higher Education Entrance Test Institute. 2022. "Top 1000 Schools by Score in the Computer-Based Written Examination, 2020 and 2021 Rankings." Accessed: 2025-01-24.

Indonesian Government. 2016. "Instruksi Presiden (INPRES) Nomor 9 Tahun 2016 tentang Revitalisasi Sekolah Menengah Kejuruan Dalam Rangka Peningkatan Kualitas dan Daya Saing Sumber Daya Manusia Indonesia." Legislation.

Jukes, Matthew, Shaher Banu Vagh, and Young-Suk Kim. 2013. "Development of Assessments of Reading Ability and Classroom Behavior - A report prepared for the World Bank." World Bank.

Kerwin, Jason T, and Rebecca L Thornton. 2021. "Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures." Review of Economics and Statistics, 103(2): 251–264.

Loyalka, Prashant, Anna Popova, Guirong Li, and Zhaoguo Shi. 2019. "Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program." American Economic Journal: Applied Economics, 11(3): 128–154.

Majerowicz, Stephanie, and Ricardo Montero. 2018. "Can teaching be taught? Experimental evidence from a teacher coaching program in Peru." Job Market Paper.

Menteri Pendidikan dan Kebudayaan. 2015. "Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 43 Tahun 2015 Tentang Uji Kompetensi Guru Atau Pendidik Lainnya dan Tenaga Kependidikan." 43.

Ministry of Education, Culture, Research, and Technology of Indonesia (MoEC). 2023. "Asesmen Nasional (AN) and Rapor Pendidikan Public Data." Accessed: January 15, 2025.

National Accreditation Board for Early Childhood, Basic, and Secondary Education (BAN-PDM). 2025. "School Accreditation Data Archive." Accessed: 2025-01-24.

National Center for Education Statistics (NCES). 2021. "2020–21 National Teacher and Principal Survey (NTPS): Teacher Questionnaire." Accessed: 2025-01-24.

Newhouse, David, and Daniel Suryadarma. 2011. "The value of vocational education: High school type and labor market outcomes in Indonesia." The World Bank Economic Review, 25(2): 296–322.

Null, Clair, Clemencia Cosentino, Swetha Sridharan, and Laura Meyer. 2017. "Policies and programs to improve secondary education in developing countries." Mathematica Policy Research.

Peugh, James L, Daniel Strotman, Meghan McGrady, Joseph Rausch, and Susmita Kashikar-Zuck. 2017. "Beyond intent to treat (ITT): a complier average causal effect (CACE) estimation primer." Journal of school psychology, 60: 7–24.

Popova, Anna, David K. Evans, Mary E. Breeding, and Violeta Arancibia. 2022. "Teacher Professional Development around the World: The Gap between Evidence and Practice." The World Bank Research Observer, 37(1): 107–136.

Pritadrajati, Dyah. 2018. "From School to Work: Does Vocational Education Improve Labour Market Outcomes? An Empirical Analysis of Indonesia." UNESCAP.

Rodriguez-Segura, Daniel, and Beth E Schueler. 2022. "Can learning be measured by phone? Evidence from Kenya." Economics of Education Review, 90: 102309.

Schaffner, Julie, Paul Glewwe, and Uttam Sharma. 2024. "Why Programs Fail: Lessons for Improving Public Service Quality from a Mixed-Methods Evaluation of an Unsuccessful Teacher Training Program in Nepal." The World Bank Economic Review, 0(0): 1–24.

Setditjen Vokasi. 2020. "Laporan Akhir Riset Evaluasi Pelaksanaan Program Pendidikan Vokasi Tahun 2020." Sekretariat Direktorat Jenderal Pendidikan Vokasi Kementerian Pendidikan dan Kebudayaan.

Stallings, Jane. 1980. "Allocated academic learning time revisited, or beyond time on task." Educational Researcher, 9(11): 11–16.

Tripepi, Giovanni, Nicholas C Chesnaye, Friedo W Dekker, Carmine Zoccali, and Kitty J Jager. 2020. "Intention to treat and per protocol analysis in clinical trials." Nephrology, 25(7): 513–517.

World Bank. 2015. "Conducting Classroom Observations: Analyzing Classroom Dynamics and Instructional Time Using the Stallings "Classroom Snapshot" Observation System."

Ye, Chenglin, Joseph Beyene, Gina Browne, and Lehana Thabane. 2014. "Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study." BMJ open, 4(6): e005362.

Yoon, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L Shapley. 2007. "Reviewing the evidence on how teacher professional development affects student achievement. issues & answers. rel 2007-no. 033." Regional Educational Laboratory Southwest (NJ1).

Yoshikawa, Hirokazu, Diana Leyva, Catherine E Snow, Ernesto Treviño, M Barata, Christina Weiland, Celia J Gomez, Lorenzo Moreno, Andrea Rolla, Nikhit D'Sa, et al. 2015. "Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes." Developmental psychology, 51(3): 309.

Zaw, Htet Thiha, Suguru Mizunoya, Dominic Richardson, Despina Karamperidou, Hiroyuki Hattori, and Monika Oledzka-Nielsen. 2021. "Teacher training and textbook distribution improve early grade reading: evidence from papua and west papua." Comparative Education Review, 65(4): 691–722.

Table 1: Correlation between vocational test teacher's scores and UKG test scores

	(1)	(2)	(3)
Overall score	0.023***		
	(0.008)		
Vocational score		0.022**	
		(0.009)	
Pedagogical score			0.016**
			(0.007)
UKG test subject FE	Yes	Yes	Yes
Province FE	Yes	Yes	Yes
Dep. Var. Mean	0.727	0.727	0.727
Dep. Var. SD	0.141	0.141	0.141
Observations	435	435	435

Notes: The table shows coefficients for standardized UKG scores. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Summary statistics by treatment status

	All (1)	Control (2)	Treated (3)	Difference (4)
Age (years)	36.20	36.11	36.30	-0.53
Male	0.67	0.71	0.63	-0.03
Has a bachelor's degree or higher	0.99	1.00	0.99	-0.00
Has a master's degree or higher	0.15	0.14	0.16	0.02
Civil servant	0.38	0.39	0.36	-0.08
Civil servant or full time employee	0.65	0.65	0.65	-0.01
Teaching experience (years)	9.73	9.66	9.80	-0.26
Salary (USD)	242.14	239.30	245.38	-28.55
Java-Bali	0.52	0.50	0.54	-0.00
Sumatera	0.24	0.23	0.26	-0.00
Kalimantan and other eastern islands	0.24	0.27	0.20	0.01
Teaches grade 10 at school	0.65	0.63	0.67	0.07
Teaches grade 11 at school	0.77	0.75	0.80	0.07
Teaches grade 12 at school	0.76	0.79	0.72	-0.09*
Total classes taught at school (grade 10-12)	4.88	4.86	4.90	-0.24
Average students per class	30.28	29.84	30.76	0.08
Program subject				
Programming/ICT/digital subjects	0.81	0.83	0.79	
Machinery/automotive subjects	0.01	0.01	0.00	
Accounting/business subjects	0.03	0.03	0.03	
Entrepreneurship subjects	0.14	0.14	0.14	
Hospitality subjects	0.00	0.00	0.00	
Fashion subjects	0.00	0.00	0.01	
Other technical subjects	0.07	0.06	0.08	
Observations	396	208	188	

Notes: Columns (2) and (3) show means for teachers assigned to the control and treatment groups, respectively. Column (4) shows the difference in means between treated and control groups, after accounting for the randomization strata. Significance levels based on robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Who applied to UR?

	Applied to UR		
	(1)	(2)	(3)
Permanent staff	0.035***	0.036***	0.034***
	(0.002)	(0.002)	(0.002)
Public school	0.048***	0.047***	0.043***
	(0.003)	(0.003)	(0.003)
Teaching certification	-0.009***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.003)
Female	-0.017***	-0.017***	-0.016***
	(0.002)	(0.002)	(0.002)
2015 Vocational subject test score	0.016***	0.017***	0.015***
	(0.001)	(0.001)	(0.001)
2015 Pedagogy test score	0.007***	0.007***	0.006***
	(0.001)	(0.001)	(0.001)
Age	-0.002***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)
Teaching experience		-0.002***	-0.001***
		(0.000)	(0.000)
Province FE	✓	✓	✓
Vocational sector FE	✓	✓	✓
University FE			✓
Dep. Var. Mean	0.080	0.080	0.080
Dep. Var. SD	0.271	0.271	0.271
Observations	109,660	109,660	109,660

Notes: The table presents coefficients from an OLS regression of a UR application dummy on pre-treatment characteristics. The sample is restricted to vocational high school teachers who took the 2015 competency test and who were between 23 and 53 years old in 2015. All regressions include province and vocational sector fixed effects. Standard errors clustered at the school level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effect estimates of treatment on UR participation

Panel A	(1) Attended UR MoEC record	(2) Trained by private firm	(3) Had any training last year
Treated	0.211***	0.169***	0.010
	(0.056)	(0.064)	(0.045)
Control mean	0.250	0.464	0.731
Observations	395	307	395
Panel B	**No. trainings last year**	**Hours in training**	**Training follow-ups**
Treated	0.449	-8.236	-0.419
	(0.437)	(15.197)	(0.351)
Control mean	2.486	64.385	0.851
Observations	395	395	395

Notes: The estimates come from OLS regression of the dependent variable on an indicator of being assigned to the treatment group. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, an urban dummy, an array of province dummies, and an array of vocational sector dummies. Estimates without controls provide similar results. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: ITT effect estimates of training on teachers' vocational knowledge and classroom practices

		Share of classroom time used for		
	(1)	(2)	(3)	(4)
Panel A	Vocational test	Lectures	Independent work	Discussion
Treated	-0.005	-0.021	0.020	0.010
	(0.016)	(0.032)	(0.032)	(0.024)
Dep. Var. Mean	0.719	0.466	0.196	0.167
Dep. Var. SD	0.144	0.237	0.240	0.168
Observations	395	311	311	311

		Share using ICT to/for		
Panel B		Cover material	Discussion	Pupil work
Treated		-0.017	0.059	0.000
		(0.051)	(0.059)	(0.061)
Dep. Var. Mean		0.739	0.359	0.423
Dep. Var. SD		0.440	0.480	0.495
Observations		395	395	395

Notes: The estimates come from OLS regression of the dependent variable on an indicator of being assigned to the treatment group. In panel A, the dependent variable in column (1) is the share of test questions answered correctly, while in columns (2) to (4) is the share of classroom time dedicated to the indicated classroom activity. In panel B, the dependent variable in columns (2) to (4) is the share of teachers saying that they use Information and Communication Technologies (ICT) to carry out the indicated classroom activity. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, an urban dummy, and an array of province dummies. All regressions include matched group fixed effects and weight control observations by the inverse of the matched control group size. Estimates without controls provide similar results. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: ITT effect estimates of training on teachers' expectations of student outcomes

	(1) Employed	(2) Wage (USD)	(3) University attendance	(4) Industry-ready
Treated	-0.006	-7.267	0.004	0.063**
	(0.019)	(5.845)	(0.018)	(0.030)
Dep. Var. Mean	0.352	136.811	0.297	0.106
Dep. Var. SD	0.208	71.166	0.206	0.308
Observations	322	310	345	388

Notes: The estimates come from OLS regression of the dependent variable on an indicator of being assigned to the treatment group. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, an urban dummy, an array of province dummies, and an array of vocational sector dummies. Estimates without controls provide similar results. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: ITT effect estimates of UR treatment on school-level outcomes

	Accreditation level		A-level accreditation		In top 1000	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.031	-0.024	-0.059	-0.056	-0.010	-0.001
	(0.040)	(0.036)	(0.048)	(0.048)	(0.014)	(0.011)
Vocational program dummies	✓	✓	✓	✓	✓	✓
In top 1000 in 2020						✓
Previous accreditation		✓				
Previous A-level accreditation				✓		
Observations	591	591	591	591	591	591

Notes: The table shows ITT estimates of the effect of being treated by UR on school-level outcomes. The dependent variable in columns (1) and (2) is the school accreditation level coded continuously, in columns (2) and (3) is an indicator of having A-level accreditation (the highest score), and in columns (5) and (6) is an indicator equal to 1 if the school ranked among the top 1000 in the university entrance examination. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Heteregeneity in treatment effects by teacher's characteristics

	Master's degree (1)	Above-median tenure (2)	Permanent employee (3)	Certified (4)	Male (5)	In Java/ Bali (6)
A. Outcome: Vocational knowledge test						
Treated	-0.003	-0.013	-0.021	-0.012	0.012	-0.046**
	(0.017)	(0.019)	(0.025)	(0.019)	(0.026)	(0.023)
Treated × group	-0.008	0.043	0.027	0.028	-0.026	0.076**
	(0.043)	(0.035)	(0.031)	(0.031)	(0.033)	(0.031)
Dep. Var. Mean	0.719	0.719	0.719	0.719	0.719	0.719
Dep. Var. SD	0.144	0.144	0.144	0.144	0.144	0.144
Observations	395	395	395	395	395	395
B. Outcome: Share of graduates employed 3 months after graduation						
Treated	-0.001	-0.008	0.023	0.001	0.037	-0.024
	(0.020)	(0.021)	(0.031)	(0.022)	(0.032)	(0.031)
Treated × group	-0.033	0.010	-0.047	-0.024	-0.067*	0.032
	(0.050)	(0.039)	(0.037)	(0.037)	(0.039)	(0.037)
Dep. Var. Mean	0.352	0.352	0.352	0.352	0.352	0.352
Dep. Var. SD	0.208	0.208	0.208	0.208	0.208	0.208
Observations	322	322	322	322	322	322
C. Outcome: Average monthly salary in first job (USD)						
Treated	-7.823	-12.028*	-8.448	-7.329	-1.664	-8.364
	(6.493)	(6.523)	(8.204)	(5.238)	(5.912)	(10.650)
Treated × group	3.783	21.288*	1.914	0.247	-8.485	1.853
	(8.518)	(12.365)	(9.644)	(11.420)	(9.562)	(10.521)
Dep. Var. Mean	136.811	136.811	136.811	136.811	136.811	136.811
Dep. Var. SD	71.166	71.166	71.166	71.166	71.166	71.166
Observations	310	310	310	310	310	310

Notes: The table shows coefficients of ITT training effects for several outcomes. Additionally, each column includes as a regressor an interaction between a dummy equal to one for the group indicated in the column heading and the assigned-to-treatment dummy. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, an urban dummy, an array of province dummies, and an array of vocational sector dummies. Robust standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendices

Appendix A Tables

Table A1: Examples of Upskilling and Reskilling courses

Target industry	Course name	Provider
Construction	Furniture and building wood finishing	Bojong Westplas LLC
	Gypsum board finishing	Petrojaya Boral Plasterboard LLC
Healthcare	Toddler care industry process	Koba Mirai
	Industrial processes to support the elderly	Koba Mirai
Hospitality	Meat processing training	Badranaya, Bandung
	Organic vegetable processing training	Prosperous Agro Mandiri, Banjarnegara
	Preparation of guest rooms / porter services	Royal Hotel Padjadjaran Bogor, The Mirah Hotel Bogor
	Manufacturing of Indonesian food	IPA, MGM Horizon
Repair of vehicles and machinery	Operation and maintenance of pneumatic equipment and systems	Festus
	Programming and operation of computer numerical control machines (CNC)	Siemens LLC
	Light vehicle chassis maintenance	Toyota Delta More Medan
	Engine management system	Toyota Delta More Medan
Software and design services	3D animation	Kayon Tunggal Ika LLC
	Computer graphics	Publisher Diamond Pariwara
	Publication design	Creative Media E&E Studio
	Photography	CPC Photo Design
	Creative digital marketing with adobe creative cloud	Adobe
	Blouse manufacturing process	Creative Creative Ditali LLC
	Network and communications	MicroTik

Notes: course names translated by the authors.

Table A2: Schedule and training partners for selected UR courses

UR 2020 Courses	OJT training partners	Schedule and time allocation		
A. BBPPMPV BOE Batch 1				
Programming and operation of CNC machines	PT. INKA, PT.PAL		Hours/Days	Dates
		Online training	65/12	21 Sep-3 Oct
Maintenance of injection system light vehicles	Auto 2000 Malang and Surabaya; Nissan Surabaya; Nasmoco Solo; Borobudur Otomobil Yogyakarta	In-person training	102/12	19-31 Oct
Topographic mapping	PT. Rasicipta Consultama, PT. Onward Bladgoud	On-the-job training	60/6	02-07 Nov
Gypsum board finishing	PT. Petrojaya Boral Plasterboard	Certification	40/4	09-12 Nov
Smartphone optimization	Service Center Samsung Electronics Indonesia	total	267/34	
B. BBPPMPV BOE Batch 3				
Steel plates welding techniques	PT. INKA, PT. PAL		Hours/Days	Dates
Light vehicle spooring and suspension system	PT Malintra	Online training	65/12	19-31 Oct
		In-person training	102/12	16-28 Nov
Using 2D AutoCAD	PT. Tiga Dinamika Solusi Indonesia	On-the-job training	60/6	30 Nov-05 Dec
Furniture and building wood finishing	PT. Natania Furniture Singosari, PT Propan Raya Industrial Coating Chemical	Certification	40/4	07-10 Dec
Building plumbing installations	PT. Bojong Wesplas	total	267/34	

Notes: This table combines and adapts information from BBPPMPV BOE (2020), page 5 (OJT partner), page 11 (time allocation), and page 26 (schedule).

Table A3: Curriculum for UR training in Network and Communications

Content	Training Modules (Contact Hours)
Policies	MoEC policy (2), Upskilling and Reskilling policy (2)
MTCNA	Introduction (2), DHCP Server and Client, ARP (2), Bridging, Wireless Bridging (4), Foundations of Routing (4), Wireless (4), Firewall (4), Quality of Service (4)
	Mikrotik Certified Network Associate (MTCNA) Certification test (3)
MTCRE	Static Routing (5), Point to Point Address (2), VPN (2), Open Shortest Path First (9)
	MikroTik Certified Routing Engineer (MTCRE) certification test (3)
Fiber optics	Intro to Fiber Optic (3), Fiber Optic Network Design FTTX FFTH (5), Fiber Optics Cable Installations, Optical Distribution Panel Adapter, Optical Terminal Box (4), Fusion Splicer and Mechanical Fiber Optics Termination (8), Damping Measurement (2), Fiber Optic Cables Implementation for Internet access using Mikrotik and SFP Module (2), Troubleshooting (3)
	Fiber Optics test (5)

Notes: This table presents the curriculum delivered during the offline training organized by BBPPMPV KPTK (a government training provider) in two batches. Batch 1 took place between 26 Oct-06 Nov 2020 at the BBPPMPV KPTK building, and Batch 2 took place between 9-20 November in the Hotel Gammara Makassar. The curriculum table is taken from *'Laporan Singkat Program Upskilling dan Reskilling Guru Kejuruan SMK Kompetensi Keahlian Teknik Komputer Jaringan "Mikrotik dan Fiber Optik"'*, a document issued by BPPMPV KPTK Direktorat MitrasDUDI Dirjen Pendidikan Vokasi Kemdikbud. The above table is adapted from the table on page 2 of the report. Nineteen out of twenty participants Batch 1 scored 89.9 in the final exam, while one scored 83.95. Nineteen out of 20 participants in Batch 2 scored 94.1, while one scored 90.53. Participants' post-training scores are summarized in tables on pages 7 and 11 of the report.

Table A4: Distribution of applicants, budgeted slots available, and number of actual attendees by course

Sector code	Training name	(1) Total applicants	(2) Budgeted slots	(3) Attendees
A. Trainings in RCT design				
TKJ 1	Mikrotik and fiber optics	1355	20	60
RPL 2	Java programming	357	150	206
RPL 4	Database management	280	150	146
GEO 1	Topographic mapping	87	40	33
SIJ 2	Internet of things	79	20	15
ANI 3	2D animation	77	25	16
B. Trainings in PSM design				
AK 1	Accounting processing training	1575	80	75
OTO 2	Maintenance of injection system light vehicles	1383	40	96
OTKP 1	Training for administrative staff	1348	80	76
ANI 1	Creative digital marketing with adobe creative cloud	1341	100	39
OTO 3	Automotive mechanic junior, light vehicle chassis maintenance	595	60	57
HOT 1	Industrial process preparation of guest rooms	586	80	66
SAM 1	Smartphone optimization and smartphone troubleshooting	579	40	47
RPL 6	Android programming	571	28	67
OTO 1	Light vehicle suspension and spooring balancing system	532	40	53
CNC 1	CNC machine programming and operation	528	92	99
LIS 1	Center of excellence training for vocational school teachers	525	108	9
BUS 1	Blouse manufacturing process	507	80	79
TEI 1	Operation and maintenance of pneumatic equipment and systems	451	40	33
BOG 1	Continental and oriental food manufacturing industry processes	445	80	70
AGRI 3	Fruit and vegetable processing training	422	14	18

Note: Panel A shows a list of all the sectors slotted for the RCT design. Panel B shows the list of the 15 largest classes in the PSM design by the number of applicants. Column (2) shows the number of student slots *initially* budgeted by MoEC. In some cases, these slots were expanded by MoEC at a later date.

Table A5: Share of correct answers to vocational knowledge questions by question and program

Question number	(1) All	(2) GEO1	(3) ANI3	(4) TKJ1	(5) SIJ2	(6) RPL2	(7) RPL4
1	0.86	0.93	0.67	0.90	0.78	0.79	0.93
2	0.57	0.57	0.29	0.61	0.69	0.76	0.36
3	0.87	0.50	0.95	0.97	0.88	0.91	0.82
4	0.82	0.87	0.86	0.90	0.78	0.82	0.74
5	0.87	0.87	0.76	0.98	0.66	0.80	0.89
6	0.69	0.33	0.14	0.74	0.94	0.79	0.65
7	0.75	0.80	0.81	0.84	0.66	0.73	0.68
8	0.81	0.73	0.76	0.82	0.69	0.87	0.79
9	0.86	0.77	0.81	0.78	0.88	0.87	0.95
10	0.28	0.20	0.14	0.076	0.94	0.31	0.31
Answered all True	32	5	2	4	0	7	14
Share all True	0.070	0.17	0.095	0.034	0	0.053	0.12
Mean score	0.72	0.66	0.62	0.76	0.79	0.76	0.71
Observations	454	30	21	119	32	131	121

Notes: The table the share of correct answers in the vocational knowledge test by question and program.

Table A6: Correlations between national exam scores and teachers' reported expectations of their schools' graduates

	Dep. Var.: 2019 Grade 12 National Exam				
	Average	**Math**	**Indonesian**	**English**	**Vocational**
	(1)	(2)	(3)	(4)	(5)
A. Share of 2019 graduates in employment three months after graduation					
Share employed	1.86***	1.00*	1.62**	1.68***	3.15***
	(0.45)	(0.52)	(0.61)	(0.47)	(0.65)
Dep. Var. Mean	47.70	35.99	67.06	42.47	45.27
Dep. Var. SD	7.47	7.72	8.77	8.09	7.50
R2	0.39	0.33	0.52	0.32	0.27
Observations	1442	1442	1442	1442	1442
B. Average monthly salary of 2019 graduates in their first job (USD)					
Salary (USD)	0.0047*	0.0068**	0.0058*	0.0094**	-0.0033
	(0.0026)	(0.0026)	(0.0030)	(0.0035)	(0.0024)
Dep. Var. Mean	47.91	36.12	67.34	42.70	45.48
Dep. Var. SD	7.44	7.67	8.69	8.06	7.54
R2	0.38	0.33	0.52	0.31	0.26
Observations	1385	1385	1385	1385	1385
C. Share of 2019 graduates continuing to university after graduation					
Share in university	7.31***	6.36***	7.69***	9.98***	5.22***
	(1.53)	(1.66)	(1.48)	(1.89)	(1.55)
Dep. Var. Mean	47.71	35.98	67.11	42.52	45.25
Dep. Var. SD	7.45	7.76	8.75	8.05	7.47
R2	0.42	0.35	0.55	0.36	0.28
Observations	1518	1518	1518	1518	1518
D. Share of 2019 graduates working in vocational sector after graduation					
Share in vocational work	2.39***	1.12*	2.00**	2.77***	3.66***
	(0.77)	(0.61)	(0.93)	(0.85)	(0.98)
Dep. Var. Mean	47.67	35.91	67.09	42.44	45.21
Dep. Var. SD	7.36	7.62	8.59	7.96	7.45
R2	0.39	0.33	0.52	0.32	0.27
Observations	1450	1450	1450	1450	1450
E. Subjective assessment whether 2019 graduates are industry-ready					
Industry ready	0.89***	0.71*	0.90***	1.04***	0.91**
	(0.30)	(0.36)	(0.31)	(0.31)	(0.36)
Dep. Var. Mean	47.80	36.05	67.17	42.65	45.31
Dep. Var. SD	7.51	7.82	8.77	8.16	7.51
R2	0.40	0.33	0.53	0.32	0.27
Observations	1625	1625	1625	1625	1625

Note: Coefficients from regressions of 2019 national examination scores on teachers' reported expectations of their graduates' outcomes. All regressions include an array of province dummies. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Summary statistics of UR applicants by match status with UKG test score data

	Age in 2020	In Java island	No. applications	Share of applicants
	(1)	(2)	(3)	(4)
Matched	38.93	0.56	1.54	0.67
	(0.05)	(0.00)	(0.01)	(0.00)
Not matched	31.20	0.52	1.62	0.33
	(0.07)	(0.01)	(0.02)	(0.00)
Number of applicants	18,448			

Notes: The table shows the characteristics of UR applicants by match status with the 2015 UKG data.

Table A8: Summary statistics by treatment status, PSM sample

	All	Control	Treated	Difference
	(1)	(2)	(3)	(4)
Male	0.50	0.47	0.54	-0.00
Age (years)	40.11	40.43	39.75	-0.49
Has a bachelor's degree or higher	1.00	1.00	1.00	0.00
Has a master's degree or higher	0.19	0.19	0.19	0.03
Civil servant	0.62	0.64	0.58	-0.00
Civil servant or full time employee	0.77	0.80	0.74	-0.02
Teaching experience (years)	13.20	13.31	13.08	-0.34
Salary (USD)	326.58	329.40	323.30	13.06
Java-Bali	0.59	0.59	0.59	-0.00
Sumatera	0.22	0.22	0.22	0.04
Kalimantan and other eastern islands	0.19	0.19	0.18	-0.04*
Teaches grade 10 at school	0.56	0.59	0.53	-0.01
Teaches grade 11 at school	0.80	0.80	0.80	-0.01
Teaches grade 12 at school	0.83	0.82	0.84	-0.01
Total classes taught at school (grade 10-12)	4.48	4.65	4.28	-0.46***
Average students per class	31.06	30.99	31.14	0.01
Program subject				
Programming/ICT/digital subjects	0.21	0.23	0.20	
Machinery/automotive subjects	0.10	0.08	0.12	
Accounting/business subjects	0.14	0.14	0.15	
Entrepreneurship subjects	0.13	0.15	0.11	
Hospitality subjects	0.13	0.14	0.10	
Fashion subjects	0.10	0.11	0.09	
Other technical subjects	0.24	0.22	0.26	
Observations	1,154	616	538	

Notes: Columns (2) and (3) show means for matched controls and UR attendees, respectively. Column (4) shows the difference in means between treated and control groups, net of matched groups fixed effects. Significance levels based on robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: UR training characteristics as reported by survey participants

	Mean (1)	Min (2)	Max (3)
Training duration (in hours)	274.14	10.00	1,380.00
Trained in government facility (Balai Besar)	0.88	0.00	1.00
Trained by private sector firm	0.94	0.00	1.00
Activity:			
Discussion	0.84	0.00	1.00
Teaching practice	0.28	0.00	1.00
Practice with computer	0.60	0.00	1.00
Internship at industry	0.20	0.00	1.00
Facilities received:			
Craft material	0.83	0.00	1.00
Lesson plan/video	0.49	0.00	1.00
Benefits from training:			
Industry certification	0.66	0.00	1.00
Knowledge and skill increase	0.59	0.00	1.00
Most helpful training component:			
On the job training	0.29	0.00	1.00
Training material	0.24	0.00	1.00
Incorporated content to day-to-day teaching	0.87	0.00	1.00
Share material with other teachers	0.78	0.00	1.00
Knew all material pre-training	0.21	0.00	1.00
Knew some material pre-training	0.64	0.00	1.00
Has taught material before training	0.82	0.00	1.00
Subjective score for training	8.84	5.00	10.00
Recommend graduates to training	0.76	0.00	1.00
Observations	767		

Notes: The table includes all UR training participants with valid responses for all the listed training characteristics.

Table A10: UR effect estimates on vocational knowledge and classroom practices

		Share of classroom time used for		
	(1) Vocational test	(2) Lectures	(3) Independent work	(4) Discussion
A. ITT excluding worst-compliance sector (RPL2)				
Treated	0.007	0.024	0.000	-0.007
	(0.022)	(0.044)	(0.045)	(0.031)
Dep. Var. Mean	0.720	0.466	0.191	0.170
Dep. Var. SD	0.144	0.240	0.242	0.164
Observations	272	207	207	207
B. Per-protocol				
Treated	0.036*	-0.011	0.034	0.003
	(0.020)	(0.045)	(0.046)	(0.029)
Dep. Var. Mean	0.728	0.463	0.203	0.164
Dep. Var. SD	0.134	0.242	0.246	0.166
Observations	259	195	195	195
C. 2SLS				
Attended	-0.022	-0.115	0.106	0.054
	(0.075)	(0.164)	(0.165)	(0.120)
Dep. Var. Mean	0.719	0.466	0.196	0.167
Dep. Var. SD	0.144	0.237	0.240	0.168
Observations	395	311	311	311
D. 2SLS excluding worst-compliance sector (RPL2)				
Attended	0.020	0.073	0.001	-0.020
	(0.062)	(0.125)	(0.126)	(0.087)
Dep. Var. Mean	0.720	0.466	0.191	0.170
Dep. Var. SD	0.144	0.240	0.242	0.164
Observations	272	207	207	207

Notes: The table shows the effect estimates of UR on vocational knowledge and classroom practice outcomes. In column (1), the dependent variable is the share of vocational knowledge questions answered correctly, while columns (2) to (4) use as the dependent variable the share of classroom time spent on the activity indicated in the column heading. Panel A presents ITT estimates that exclude the sector with the worst compliance with treatment assignment. Panel B restricts the sample to people who complied with the treatment assignment, i.e. attended training when assigned to the treatment group and did not attend when assigned to the control. Panel C and D show 2SLS estimates for the whole sample and excluding the worst compliance sector, respectively. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, an urban dummy, and an array of province and vocational sector dummies. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Likelihood of UR attendance by treatment assignment status

	Treated (1)	Control (2)	Treated (3)	Control (4)	Treated (5)	Control (6)
Female	0.125	0.013	0.094	0.019	0.095	0.019
	(0.079)	(0.070)	(0.080)	(0.076)	(0.080)	(0.076)
Teaching experience (years)	0.005	0.012	0.004	0.006	0.004	0.006
	(0.012)	(0.008)	(0.011)	(0.009)	(0.011)	(0.009)
Permanent employee	0.128	0.126**	0.126	0.109	0.124	0.109
	(0.090)	(0.063)	(0.088)	(0.068)	(0.090)	(0.069)
Age	-0.003	-0.012*	0.001	-0.007	0.001	-0.007
	(0.009)	(0.007)	(0.009)	(0.008)	(0.009)	(0.008)
Urban area	0.190*	0.088	0.220**	0.051	0.220**	0.051
	(0.099)	(0.081)	(0.098)	(0.086)	(0.099)	(0.086)
Vocational UKG score			-0.073*	0.007	-0.074*	0.007
			(0.042)	(0.035)	(0.043)	(0.035)
Pedagogical UKG score					0.007	-0.002
					(0.041)	(0.033)
Vocational sector FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	0.549	0.250	0.545	0.267	0.545	0.267
Dep. Var. SD	0.499	0.434	0.499	0.444	0.499	0.444
Observations	182	200	178	176	178	176

Notes: The table shows OLS results of an indicator that the teacher complied with the treatment assignment. UKG test scores are standardized. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: UR effect estimates on ICT use in the classroom

	Share using ICT to/for		
	(1)	(2)	(3)
	Cover material	Discussion	Pupil work
A. ITT excluding worst-compliance sector (RPL2)			
Treated	0.041	0.025	0.037
	(0.072)	(0.075)	(0.083)
Dep. Var. Mean	0.713	0.324	0.397
Dep. Var. SD	0.453	0.469	0.490
Observations	272	272	272
B. Per-protocol			
Treated	0.041	0.154*	0.063
	(0.076)	(0.081)	(0.084)
Dep. Var. Mean	0.703	0.351	0.378
Dep. Var. SD	0.458	0.478	0.486
Observations	259	259	259
C. 2SLS			
Attended	-0.082	0.278	0.001
	(0.229)	(0.262)	(0.273)
Dep. Var. Mean	0.739	0.359	0.423
Dep. Var. SD	0.440	0.480	0.495
Observations	395	395	395
D. 2SLS excluding worst-compliance sector (RPL2)			
Attended	0.124	0.076	0.112
	(0.202)	(0.209)	(0.233)
Dep. Var. Mean	0.713	0.324	0.397
Dep. Var. SD	0.453	0.469	0.490
Observations	272	272	272

Notes: The table shows the effect estimates of UR on Information and Communication Technologies (ICT) use in the classroom. Each column uses as the dependent variable an indicator of whether the teacher uses ICT for conducting the classroom activity indicated in the column header. Panel A presents ITT estimates that exclude the sector with the worst compliance with treatment assignment. Panel B restricts the sample to people who complied with the treatment assignment, i.e. attended training when assigned to the treatment group and did not attend when assigned to the control. Panel C and D show 2SLS estimates for the whole sample and excluding the worst compliance sector, respectively. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, an urban dummy, and an array of province and vocational sector dummies. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: UR effect estimates on teachers' expectations of student outcomes

	(1) Employed	(2) Wage (USD)	(3) University attendance	(4) Industry-ready
A. ITT excluding worst-compliance sector (RPL2)				
Treated	-0.015	-12.783	-0.005	0.079*
	(0.025)	(8.446)	(0.017)	(0.041)
Dep. Var. Mean	0.349	140.917	0.286	0.116
Dep. Var. SD	0.214	70.337	0.203	0.321
Observations	218	208	239	267
B. Per-protocol				
Treated	0.015	-8.297	0.003	0.110**
	(0.026)	(10.035)	(0.022)	(0.046)
Dep. Var. Mean	0.355	137.548	0.303	0.110
Dep. Var. SD	0.208	75.437	0.203	0.314
Observations	206	201	220	254
C. 2SLS				
Attended	-0.040	-43.346	0.024	0.295**
	(0.117)	(34.917)	(0.095)	(0.142)
Dep. Var. Mean	0.352	136.811	0.297	0.106
Dep. Var. SD	0.208	71.166	0.206	0.308
Observations	322	310	345	388
D. 2SLS excluding worst-compliance sector (RPL2)				
Attended	-0.050	-40.767	-0.016	0.223**
	(0.077)	(26.136)	(0.050)	(0.105)
Dep. Var. Mean	0.349	140.917	0.286	0.116
Dep. Var. SD	0.214	70.337	0.203	0.321
Observations	218	208	239	267

Notes: The table shows UR effect estimates on teachers' expectations of student labor market outcomes. Each column uses as the dependent variable teacher's estimate of the outcome indicated in the column header for the cohort of students graduating in 2021 (post-UR). Panel A presents ITT estimates that exclude the sector with the worst compliance with treatment assignment. Panel B restricts the sample to people who complied with the treatment assignment, i.e. attended training when assigned to the treatment group and did not attend when assigned to the control. Panel C and D show 2SLS estimates for the whole sample and excluding the worst compliance sector, respectively. All regressions include the following covariates: gender, years of education, years of teaching experience, an indicator of being a civil servant or a full-time staff, an indicator of certification status, teachers' expectations for the cohort graduating in 2019 (pre-UR), an urban dummy, and an array of province and vocational sector dummies. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: UR effect estimates on students' National Exam (AN) results

	Literacy		Numeracy	
Panel A: ITT	(1)	(2)	(3)	(4)
Treated	-0.279	-2.331	1.400	0.273
	(2.612)	(2.687)	(1.225)	(1.153)
Dep. Var. Mean	57.906	58.049	48.942	48.970
Dep. Var. SD	11.820	11.819	7.983	7.935
Observations	1270	1238	1273	1269
Panel B: 2SLS				
Attended	-1.523	-17.837	6.030	1.532
	(15.244)	(50.085)	(8.357)	(6.694)
Dep. Var. Mean	57.906	58.049	48.942	48.970
Dep. Var. SD	11.820	11.819	7.983	7.935
Observations	1270	1238	1273	1269

Notes: The table shows UR effect estimates on student test scores in the 2021 National Assessment using individual-level data. All regressions control for sector fixed effects. In addition, columns (2) and (4) control for a male dummy, student SES and school SES. Standard errors clustered at the school level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

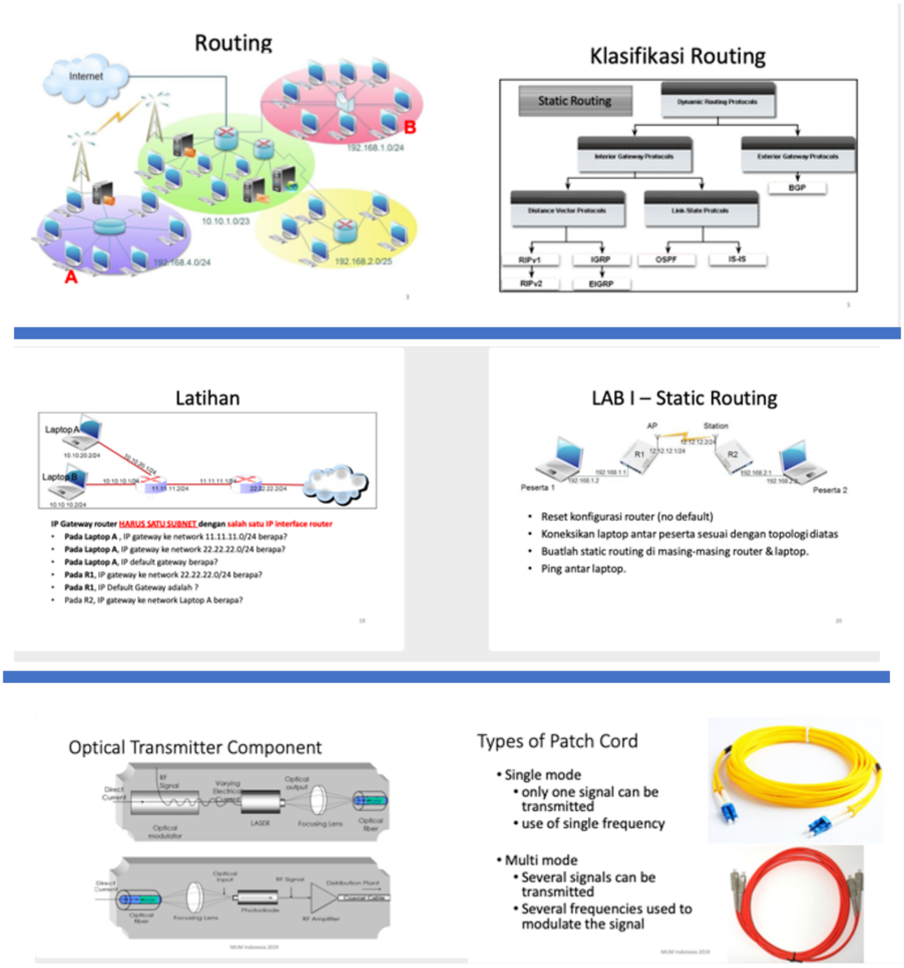
Table A15: PSM UR effect estimates

		Share of classroom time used for		
	(1)	(2)	(3)	(4)
A. Classroom practices		Lectures	Independent work	Discussion
Attended		-0.027	0.027	0.004
		(0.016)	(0.017)	(0.011)
Dep. Var. Mean		0.431	0.241	0.162
Dep. Var. SD		0.244	0.260	0.163
Observations		844	844	844
		Share using ICT to/for		
B. ICT use		Cover material	Discussion	Pupil work
Attended		0.067***	0.093***	0.022
		(0.023)	(0.029)	(0.030)
Dep. Var. Mean		0.812	0.413	0.468
Dep. Var. SD		0.391	0.493	0.499
Observations		985	985	985
	Expectations of students' outcomes			
C. Teachers' expectations	Employed	Wage (USD)	University attendance	Industry-ready
Attended	0.020*	-1.599	-0.017**	0.046***
	(0.012)	(2.899)	(0.008)	(0.016)
Dep. Var. Mean	0.385	140.350	0.232	0.080
Dep. Var. SD	0.217	66.611	0.180	0.272
Observations	702	694	775	956

Notes: The table shows Propensity Score Matching UR effect estimates on various outcomes. UR training attendees were matched to control teachers who applied to the same training. All regressions matched-group and province fixed effects, and they weight control observations by the inverse of the number of teachers in the control group. Panel A uses as dependent variable the share of classroom time spent on the indicated activities. Panel B uses as dependent variable an indicator of whether the teachers use Information and Communication Technologies (ICT) to conduct the indicated classroom activity. The dependent variables in Panel C are teachers' estimates of labor market outcomes for the student cohort graduating in 2021. All regressions control for an urban dummy and province fixed effects. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B Figures

Figure B1: Snippets of UR 2020 Training Materials



Note: Snippets of training materials from the Network and Communication training, covering modules on routing and fiber optics. Snippets are reproduced from the BPPMPV KPTK training report.

Figure B2: Snapshots of UR 2020 Training Activities

Note: Photos taken during offline training of one of the Upskilling Reskilling 2020 training. Top left: trainers and MoEC officials on a panel in front of a backdrop with the UR course title name, dates, and location. Top right and bottom left: teachers participated in hands-on activities with network equipments during the training. Bottom left: participating teachers working on individual laptops as part of the training session. Photographs are reproduced from BPPMPV KPTK training report.

Figure B3: Distribution of 2015 Teacher Competency Exam scores for survey respondents

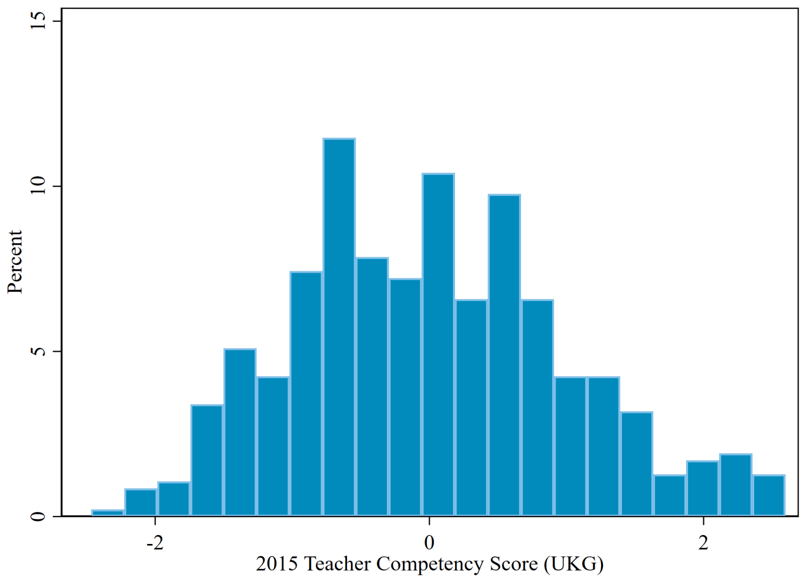
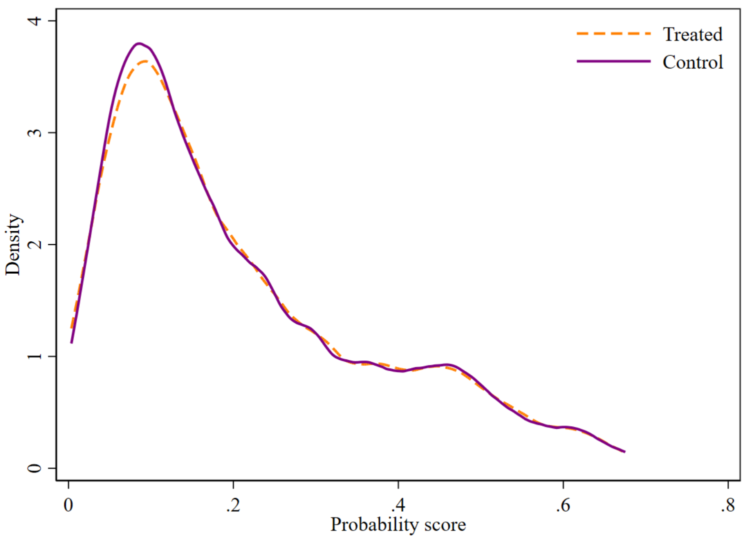


Figure B4: UR: propensity score by treatment status



Note: The figure shows estimates of the propensity score for UR attendees (treated) and people in the control group. The figure combines the scores for all trainings. The distribution of the control group is weighted by the inverse of the number of units in the control group.

Appendix C Propensity Score Matching and Sample Selection

We supplemented the RCT design with a sample of nearly 1,200 teachers who were selected using Propensity Score Matching. We used this second strategy to improve precision and to compare its results with the RCT. Our rationale was that if we obtained similar results under both strategies, this would bolster the –potentially imprecise– results of the RCT.

We included a total of 48 courses in this supplementary design. We matched each of the attendees to these 48 courses to control applicants using a Propensity Score:

$$P(T_i = 1 | \text{ subject} = j) = X_i \beta_j \tag{C1}$$

we calculate the propensity score by OLS using a series of pre-treatment characteristics available in the 2015 Teacher Competency Exam database. These are: years of education, years of teaching experience, gender, whether the teacher resides in Java island, whether the teacher was certified,[20] school type (public/private), a state-employee dummy, type of contract (permanent/temporary), field of specialization (care services, construction, creative economy, hospitality, machinery, other), and the standardized test score in the 2015 competency exam.

We estimated (C1) on the set of attendees and applicants with 2015 test score data. We ran a separate regression for each of the 48 training courses. We computed the propensity scores separately by subject because the selection of applicants was fairly decentralized, and the selection procedures could vary by course. For each subject, we included all attendees and, as potential controls, we used all people who applied for admission to that subject. Because people often applied to several trainings, this means that the same individual can appear in the control pool for several subjects. Figure B4 shows estimates of the propensity score by treatment group for all the trainings in the sample. Note that the estimated scores for the control group match quite closely the UR attendee's scores (treated group).

We matched attendees to controls using the four nearest neighbors with replacement with a caliper of 0.05; that is, for each treated individual we matched up to four controls as long as the difference between the treated and control propensity scores was within 5 percentage points. We slated all the attendees and their four matched controls for the survey. To be included in the results, we had to successfully survey the UR attendee and at least one of their matched controls.

[20] Teachers can get certified on their teaching fields, which unlocks salary supplements.