



ADB Working Paper Series

**LEVERAGING SCHOOL PRINCIPALS
TO ADDRESS LEARNING LOSS
IN INDONESIA THROUGH GROUP
AND INDIVIDUAL TARGETING**

Dewi Susanti, Gautam Anand,
Firda Arianti Arifin, and
Patrya Pratama

No. 1392
June 2023

Asian Development Bank Institute

Dewi Susanti is Senior Director of Research, and Gautam Anand is Director of Research, at Global School Leaders. Firda Arianti Arifin is Project Officer, and Patrya Pratama is Executive Director at Inspirasi Foundation.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Discussion papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. Some working papers may develop into other forms of publication.

Suggested citation:

Susanti, D., G. Anand, F. A. Arifin, and P. Pratama. 2023. Leveraging School Principals to Address Learning Loss in Indonesia through Group and Individual Targeting. ADBI Working Paper 1392. Tokyo: Asian Development Bank Institute. Available: <https://doi.org/10.56506/PXOO8818>

Please contact the authors for information about this paper.

Email: dewi@globalschoolleaders.org

We thank our partners at the Ministry of Education, Culture, Research, and Technology, the Karawang District Education Department, and the principals, teachers, and students from participating primary schools. We thank Noam Angrist, Colin Crossley, Claire Cullen, and Saipramod Bathena for their generous feedback to the initial study design and sharing of training materials and research instruments. We are grateful to Daniel C. Suryadarma for his advice and comments from the study design to the final draft of the paper. We appreciate the comments we received from Amrita Ahuja, Valmik Ahuja, Avni Gupta-Kagan, Azad Oommen, Sameer Sampat, Cici Tri Wanita, and participants of the 2022 Conference on the Quality of Basic Education in Southeast Asia, organized by ADBI and ISEAS. We acknowledge financial support from the Asian Development Bank Institute to conduct the study. This study was registered at the University of Gadjah Mada (KE/UGM/020/EC/2022) with an approval date of 8 July 2022. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the Global School Leaders and Inspirasi Foundation.

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

© 2023 Asian Development Bank Institute

Abstract

In Indonesia, COVID-19 pandemic-induced school closures led to a significant loss of learning among students, necessitating remedial learning programs. This study thus aimed to explore whether differentiated remedial teaching can improve the foundational numeracy skills of students and if the improvements would be better in schools with added individual tutoring. It also aimed to understand whether additional training of school principals would strengthen the results. After implementing four interventions in 25 primary schools to 1,545 students for eight weeks, we found that all interventions significantly improved the foundational numeracy skills of students, students in schools with added individual tutoring did better than those without the added tutoring, and teachers in schools with trained principals were more punctual in implementing the intervention.

Keywords: learning loss, group remedial, individual tutoring, school principal support

JEL Classification: I21

Contents

1.	INTRODUCTION	1
2.	RESEARCH DESIGN	3
2.1	Research Questions and Study Design.....	3
2.2	Research Method and Data Collection.....	4
2.3	School Selections and Assignments	5
3.	IMPLEMENTATION.....	6
4.	FINDINGS	7
5.	CONCLUSIONS AND RECOMMENDATIONS	11
	REFERENCES	13
	ANNEXURES	16

1. INTRODUCTION

The COVID-19 pandemic paralyzed education systems worldwide, forcing over 1.6 billion learners out of classrooms at its peak (UNESCO 2022). The World Bank estimated that school closures could cost up to \$10 trillion in net present value (Azevedo et al. 2020). A review of research on learning loss due to the pandemic found that the loss of learning has been much higher among students with lower socioeconomic status in any context, even if there has been no learning loss in a country (Moscoviz and Evans 2022).

In Indonesia, schools were closed for most of its 62 million students over 19 months. Some estimated that the closures could lead to a reduction of between 21 and 35 points in Programme for International Student Assessment (PISA) reading scores (UNICEF 2021; Afkar and Yarrow 2021), and translate into a reduction in lifetime earnings of up to USD359 billion in present value (Afkari and Yarrow 2021).

If not addressed, the impact of the pandemic in Indonesia may further contribute to a downward trend in student outcomes. Between 2000 and 2018, Indonesia consistently ranked amongst the lowest ten countries in PISA assessments and experienced a hump-shaped trajectory for language, math, and science (Beatty et al. 2021). Another longitudinal study that spanned from 2000 to 2014 using nearly nationally representative data also found that Indonesia's student outcomes in mathematics had declined (Beatty et al. 2021).

The Indonesian Ministry of Education, Culture, Research, and Technology (MoECRT) has implemented several reforms¹ to address learning loss. They adapted the "Teaching at the Right Level" (TaRL) approach by assessing student learning and modifying the curriculum and teaching to students' level. TaRL has been rigorously evaluated in various contexts using different approaches, ranging from dedicated time during school hours to in-school learning or holiday camps, and out-of-school community groups led by trained facilitators (Banerjee et al. 2007, 2016, 2017; Duflo et al. 2008, 2011, 2020). In the context of addressing learning loss due to the pandemic, a government-led intervention in India where volunteers conducted foundational learning in schools and other community spaces also led to a rapid learning recovery (Singh, Romero and Muralidharan 2022).

TaRL has also inspired interventions that target foundational learning at the individual student level using technology, employing both AI-powered adaptation (Muralidharan, Singh, and Ganimian 2017) and low-tech phone-based tutoring (Angrist, Bergman, and Matsheng 2022; Crawford et al. 2021). During the pandemic, Youth Impact conducted an experiment in Botswana that combined weekly SMS text messages containing a few basic numeracy problems with a 20-minute numeracy phone tutoring session. Students benefited significantly from the intervention, with innumeracy being reduced by 52% compared to the control group (Angrist et al. 2020).

In 2020, Youth Impact replicated this intervention in five other countries, including India (Angrist et al. 2023). A unique feature of the intervention in India was the involvement of school principals in the delivery of the program through teachers. Evidence that strong school leadership improves student outcomes has been established in the United States (Grissom, Egalite, and Lindsay 2021) and in low- and middle-income countries (LMICs) through the World Management Survey (Scur et al. 2021). However, the evidence is mixed on which area of school leadership training improves

¹ <https://merdekabelajar.kemdikbud.go.id/en/utama>.

learning outcomes in LMICs (De Barros et al. 2022; Muralidharan and Singh 2020; Romero et al. 2021; Cilliers and Habyarimana 2021; Beg et al. 2021), which calls for more studies.

Initial findings from Youth Impact's study in India showed that in addition to significantly improved student learning outcomes (Angrist et al. 2023). A follow-up qualitative study to complement Youth Impact's study in India shows that including school principals in the training for teachers improves support for teachers, enabling teachers to adapt the approach to other subjects, and improving the potential for sustaining the intervention. The study also found that leveraging the role of school principals can potentially reduce the role of NGOs in the intervention. But school principals need specific training on utilizing student learning outcomes data to provide instructional leadership. In addition, some school principals and teachers thought that the approach could not be sustained once schools have fully opened to students, as they would not have the time to make individual calls to all of their students.

Given the massive scale of the pandemic's impact on learning loss, with students from disadvantaged backgrounds having lost learning much more significantly than others, this study aimed to answer the following research questions: (1) Could teachers provided with a short training on conducting foundational numeracy assessment and differentiated teaching implement the approach? (2) Would student numeracy outcomes improve in classrooms that implement differentiated teaching? (3) Would student numeracy improvements be better in schools where teachers conduct additional individual tutoring for students facing the most challenges?

In addition, all of the aforementioned group foundational learning and individual tutoring interventions, while effective, targeted teachers as the key lever of change. However, many LMICs struggle to conduct and support teacher professional development even during normal times. An often underinvested area with potential high return is training school leaders to mentor teachers. Hence, this study posed an additional question: (4) Would additional training of school principals strengthen the results of an education intervention delivered through teachers?

Our research design is informed by the Indonesian Ministry of Education, Culture, Research, and Technology (MoECRT) broader reforms, and through a series of consultations with MoECRT staff. From 2020 until the end of 2022, the MoECRT rolled out 22 reform episodes, with three of them being relevant for this study. The first one was the minimum student competence assessment (*Asesmen Kompetensi Minimum* or AKM) to help teachers diagnose Grade 5 students' literacy, numeracy, and behavioral characteristics. The second was a platform for teachers to get access to references and best practices, including on how to differentiate teaching (*Platform Merdeka Mengajar* or PMM). Through a third reform, the MoECRT targets school and teacher leaders (*Guru Penggerak*) to become instructional leaders, foster communities of professional development practices, and establish stronger relationships with parents and broader communities. The MoECRT recognizes that most schools face challenges in weaving the TaRL-inspired reforms together.

In 2022, we implemented a pilot in 25 public primary schools in the district of Karawang, West Java, Indonesia. The pilot compared four interventions. All groups trained Grade 3 and 4 teachers on conducting numeracy assessment, grouping students based on their numeracy levels, targeting teaching based on the relevant mathematical operations, and reassessing students frequently to adjust their levels (basic intervention).

To answer our research questions, we directly compared two approaches and four intervention groups. In all schools, we implemented a group foundational learning based on student numeracy levels (basic model, Group A). Specifically, we trained Grade 3 and 4 teachers to conduct student numeracy assessment, target learning to students' level, and reassess frequently to adjust student leveling. In a randomly assigned half of the schools, in addition to the basic model, every two weeks teachers were asked to deliver a targeted individual tutoring to ten students who struggled the most in foundational numeracy (basic + tutoring model, Group B). In addition, we also randomly assigned half of the schools, where we trained school principals to support teachers in implementing either basic (Group C) or basic + tutoring (Group D) models.

The pilot was implemented over eight weeks, involving 1,545 students through training 50 teachers and 23 principals². We evaluate the outcomes using quantitative data on student assessments at baseline, throughout implementation, and at endline, and circulated surveys to all principals and teachers at baseline and endline. We also conducted interviews with principals and teachers in eight schools (two from each of the intervention groups), and observed their check-in meetings.

We identified six key findings: (1) trained teachers could implement the targeted group and individual tutoring; (2) The biggest challenge for implementation was in scheduling the sessions, while the biggest barrier was teachers' belief in students' ability to progress; (3) all the interventions were effective in improving student outcomes; (4) two interventions performed the best: (a) trained teachers implementing the group and individual targeting sessions, and (b) trained teachers supported by a trained school principal implementing the group targeting sessions; (5) teachers in schools with trained principals had better compliance in inputting student outcomes data on time; and (6) in addition to instructional leadership skills, school principals need additional training on management and soft skills.

The next section details the research questions, design, method, and school selections and assignments. Section 3 describes the implementation. Section 4 presents the results on student learning outcomes followed by a brief conclusion.

2. RESEARCH DESIGN

2.1 Research Questions and Study Design

We asked four research questions: (1) Could teachers provided with a short training on conducting foundational numeracy assessment and differentiated teaching implement the approach? (2) Would student numeracy outcomes improve in classrooms that implement differentiated teaching? (3) Would student numeracy improvements be better in schools where teachers conduct added individual tutoring for students facing the most challenges? (4) Would additional training of school principals strengthen the results of an education intervention delivered through teachers?

We directly compared two approaches and four intervention groups to answer our research questions (Figure 1). In all schools, we implemented a group foundational learning based on student numeracy levels. Grade 3 and 4 teachers were trained to assess student numeracy across five levels: addition, subtraction, multiplication, division without remainders, and division with remainders. They were asked to group students into the five levels and target teaching according to these levels in additional

² Two of the school principals were in charge of two schools each.

remedial sessions every two weeks. At the end of the sessions, teachers reassessed and readjusted student leveling.

The 25 schools were randomly assigned to four intervention groups. Group A implemented the basic model described in the previous paragraph. In Group B, in addition to the basic model, every two weeks, teachers were asked to identify ten students who struggled the most and deliver a targeted individual tutoring to them based on their level. In Groups C and D, we trained school principals to support teachers in implementing either basic (Group C) or basic + tutoring (Group D) models.

To answer our first two research questions, we asked teachers to report the implementation of the basic group remedial sessions and the result of foundational numeracy assessments every two weeks. To answer the third research question, we compared two groups that implemented only the basic models (Groups A and C) with those that implemented the basic + individual tutoring models (Groups B and D). Finally, to answer the fourth research question, we compared the two groups where only teachers were trained (Groups A and B) with those where school principals were also trained in addition to teachers (Groups C and D).

Table 1: The Four Study Groups

	Teacher Training	Teacher Training + Principal Support
Group Remedial	A: Basic Model	C: Basic Model + Principal Training
Group Remedial + Individual Tutoring	B: Basic Model + Individual Tutoring	D: Basic Model + Individual Tutoring + Principal Training

Note: There were six schools each in Groups A, B, and D. Group C had 7 schools.

We hypothesized that schools with support from the trained school principals (Groups C or D) would perform better than teacher training groups only (Groups A and B), because trained school principals are more likely to hold teachers accountable. In other words, school principals' active involvement would lead to higher response rates, lower attrition rates, and greater improvement in learning outcomes for the students.

2.2 Research Method and Data Collection

We implemented quantitative and qualitative data collection. The quantitative data collection method was adapted from Youth Impact's low-tech remote individual tutoring program (Angrist et al. 2023), which utilized a dual approach to process monitoring: (1) pre- and post-assessments on student learning by phone; and (2) rapid self-reported monitoring by teachers to track key performance indicators related to implementation.

The following data were collected from schools: (1) numeracy outcomes from all students in Grades 3 and 4 at baseline and endline; (2) a brief survey with all participating Grade 3 and 4 teachers and school principals at baseline and endline; and (3) data on group and individual targeting implemented by teachers, the number of students that participated, and assessment results on a biweekly basis. In addition, data on the biweekly check-ins with teachers for the principal groups (C and D) were collected.

The qualitative study was implemented in eight purposively selected schools, two each from the four intervention groups. Two teachers from Grades 3 and 4 were interviewed from each school, making a total of 16 teachers. A research assistant observed the check-ins conducted by school principals between Weeks 6 and 7. At the end of the

intervention, the research assistant conducted semi-structured in-depth interviews with the eight school principals, 16 teachers, a school supervisor, and the district government official.

Both the quantitative and qualitative research instruments were developed in English. The instruments were translated from English to Bahasa Indonesia, piloted, and implemented in Bahasa Indonesia. The qualitative research findings were translated and transcribed back into English for analysis.

We compared the four groups across the outcome indicators on student numeracy, the group remedial sessions and individual tutoring implemented, and student outreach. We analyzed the data from pre-, post-, and bi-weekly student assessments to track the movement of students across five learning levels, gender, and the grade level of students. The data from before and after the survey with teachers and school principals were analyzed with a view to understanding issues related to learning loss in the school, and challenges that arose during the execution of the program. The process of monitoring data helped in understanding the effectiveness of the implementation of the program as it tracked student attrition as well as the punctuality of data input from the teachers.

Assigning learning levels a numerical value of 1–5, we first compared the average learning level score of each group at posttest with the overall average at the pretest and conducted tests for statistical significance. We also compared the average scores for combinations of groups. For example, a combined average of A and B (teacher training group only) was compared to C and D (teacher training + principal support). Similarly, a combined average of A and C (basic group remedial sessions only) was compared with a combined average of B and D (basic + individual tutoring). Using levels as a continuous variable, we also conducted regression analysis utilizing an ordinary least squares method to control for gender and grade.

We conducted an analysis of the in-depth interviews with teachers and school principals and observation notes from school visits to identify how the outcome indicators had been achieved, and to collect feedback and concerns about the implementation. The findings from the qualitative study were then analyzed with the process monitoring and student assessment data to identify intervention design improvements and potential pathways for scaling up.

2.3 School Selections and Assignments

The research was conducted in one subdistrict within the Karawang District, West Java Province, approximately 80 kilometers away from Jakarta, the capital city of Indonesia. The region is a relatively semi-urban area with a total of 939 elementary schools and a 97.2% enrollment rate at the primary level (Neraca Pendidikan Daerah 2021). Despite being relatively close to Jakarta, Karawang District's Human Development Index (70.94) is slightly below the provincial (72.45) and the national (72.29) averages (Neraca Pendidikan Daerah 2021). The quality of education remains a challenge, with 60% of primary education teachers not being certified and less than 40% of classroom infrastructures being in good condition (Neraca Pendidikan Daerah 2021).

In consultation with the Karawang District education authorities, we selected Rawamerta, a rural subdistrict with 27 primary schools. Teachers and school principals in Rawamerta had received no recent training from the local or central governments, and had no other education intervention program running in parallel from other NGOs that would influence the study's outcomes.

As the study started when the schools had just reopened following the pandemic, with alternate patterns of opening and closing, the initial intervention design prioritized an individual targeting through the phone as the basic model and group remedial session as the additional intervention. Prior to the implementation, two rounds of pre-piloting were conducted with the intention of testing the program design, the student assessment tool, data input protocols, and qualitative research activities. The first pre-pilot was implemented in one school that was not in the Rawamerta subdistrict. We implemented the Group D model to comprehensively test the training modules for teachers and school principals to implement both the individual tutoring and group targeting. This first pre-pilot was also aimed at figuring out how schools could manage the weekly schedules for both the individual tutoring and group targeting, which turned out to be very demanding for teachers. After consulting with the Youth Impact team, we adjusted the schedule so that teachers would alternate between group remedial and individual tutoring sessions each week. To account for lost time, we doubled the time for the sessions.

We then tried this design with the two schools participating in the second pre-pilot to test out the C and D models. Two schools in Rawamerta were selected for the pre-pilot. One was set up recently and did not have a national exam score yet, while the other was an outlier in terms of the national exam score for numeracy (55.37 compared to the average of 76.14). Following the evaluation of the two pre-pilots by officials from the MoECRT, the study was redesigned with the group remedial sessions as the main treatment, and individual tutoring only for students whose learning outcomes remained stagnant, to further reduce the demand on teachers' time.

We used five indicators from the administrative data to conduct a random assignment of the remaining 25 schools in Rawamerta into four groups: the total number of students, the student-teacher ratio, the ratio of certified teachers, the 2019 national examination average score in mathematics, and the average total national examination score. We followed a two-phase assignment process. First, we assigned the schools into two groups: teacher training (12 schools) and teacher + principal training (13 schools). Second, within these two groups, schools were again randomly assigned to two subgroups: group remedial sessions (A and C) and group remedial sessions + individual tutoring (B and D). The final assignments included six schools each in Groups A, B, and D, and seven schools in Group C, with a total of 1,545 Grade 3 and 4 students.

Eight schools, two each from the four groups, were selected for the qualitative data collection using a purposive sampling design. We used three major indicators to select the schools: the total number of students in the school, the punctuality of data input by teachers as a proxy for implementation quality, and the gender of the school principal. Our aim was thus to select a large and a small school, a better-performing and a low-performing school in terms of timely input of student learning assessment data, and a school led by a male and a female principal.

3. IMPLEMENTATION

In July 2022, Inspirasi, a nonprofit organization in Indonesia, trained all Grade 3 and 4 teachers from 25 schools across the four groups in: (1) assessing students' numeracy level; (2) how to conduct group foundational learning based on numeracy levels; (3) checking regularly for student understanding; and (4) undertaking individual tutoring (for Groups B and D). School principals from Groups C and D participated throughout the teachers' training and, in addition, had a half-day training on supporting teachers in

implementing students' numeracy assessments, conducting group remedial sessions, using data to monitor the implementation progress by teachers, and carrying out biweekly evaluation with teachers. School principals in Group D also received training on supporting teachers in conducting individual tutoring. We included a question in the baseline survey for the teachers to ask their preference for implementing individual tutoring: whether in person or by phone. Since the majority of them opted for in-person tutoring, the study did not implement phone tutoring.

Upon the completion of the training, the teachers started the pilot implementation by assessing students' competence in numeracy (pretest) and placed them in five levels: Level 1 – students are able to solve only basic addition; Level 2 – students are able to solve basic addition and subtraction; Level 3 – students are able to solve basic addition, subtraction, and multiplication; Level 4 – students are able to solve basic addition, subtraction, multiplication, and division without remainders; and Level 5 – students are able to solve basic addition, subtraction, multiplication, and division with and without remainders.

Teachers then divided students according to their levels and focused on teaching only one mathematical operation based on that group's numeracy level. At the end of every lesson, teachers conducted assessments for that specific mathematical operation to assess students' levels for the next meeting.

Teachers in Groups B and D had an additional task of identifying students whose levels had been stagnant over the two weeks, or those who seemed to struggle with the concepts relating to specific mathematical operations. They targeted a minimum of ten students every two weeks and conducted individual tutoring.

4. FINDINGS

We focused our analysis on student assessment data. Overall, student attrition due to long-term absence or school transfers was low, with only 1.4% between the pre- and posttest. We structure the presentation of our findings below to answer the research questions we posed.

First, after a brief training session, teachers were able to implement both the group remedial and individual tutoring sessions. Table 2 presents the number of students who attended group remedial and individual tutoring sessions across four intervention groups during the eight weeks of program implementation. The number of students who participated in individual tutoring fluctuated over the implementation period. This pattern was driven by a few teachers. Two teachers forgot the procedure in Week 2, and two wanted their students to catch up towards the end and they provided individual tutoring for all students.

Table 2

Groups	Number of Schools	Number of Students (Group Remedial)	Individual Tutoring – Week 2	Individual Tutoring – Week 4	Individual Tutoring – Week 6	Individual Tutoring – Week 8
A	6	376	–	–	–	–
B	6	363	115	77	91	104
C	7	449	–	–	–	–
D	6	357	128	68	70	69
Total	25	1,545	243	145	161	194

Note: One Grade 3 and one Grade 4 teacher participated from every school.

Second, our qualitative findings indicated that most teachers encountered challenges in scheduling the sessions and in believing in students' ability to progress. The most difficult part of the implementation for teachers was scheduling the group remedial and individual tutoring sessions, with the individual tutoring encroaching more into non-math classes. During the endline survey, we added some questions with a view to better understanding the variation in implementation (Table 3). The sessions were supposed to be conducted after school, but since many students go to religious schools in the afternoon, 84% of teachers had to fit in the sessions during school time. In group remedial interventions (A and C), 85% of teachers utilized math classes, compared to 66% in groups with individual tutoring interventions (B and D). The rest of the teachers mainly utilized fine arts and Bahasa Indonesia classes. A proportion of 77% of the teachers in group remedial interventions conducted the remedial classes at the same time and in the same classroom, compared to 50% of teachers in group and individual interventions.

Additional challenges relating to scheduling conflicts include school-wide activities (such as exams and competitions) and teachers having other roles and priorities. The program required teachers to use technology to input data from student assessments through online platforms, which some teachers and school principals found challenging due to limited digital literacy skills. They had to rely on other teachers for assistance to input data or access online data.

Some teachers associated students' lack of progress with laziness, a lack of mathematical talent, and/or a lack of motivation. Teachers seem to believe that not all students have the capability to learn. For example, a Grade 3 teacher said: "...Some students might be below the average IQ so it's hard for them to understand the concepts. It takes time and they easily forget the lessons..." Another Grade 3 teacher commented that "...the difficulty is that the students have different abilities. There are some who can hardly follow the program and can't level up (despite the program)..."

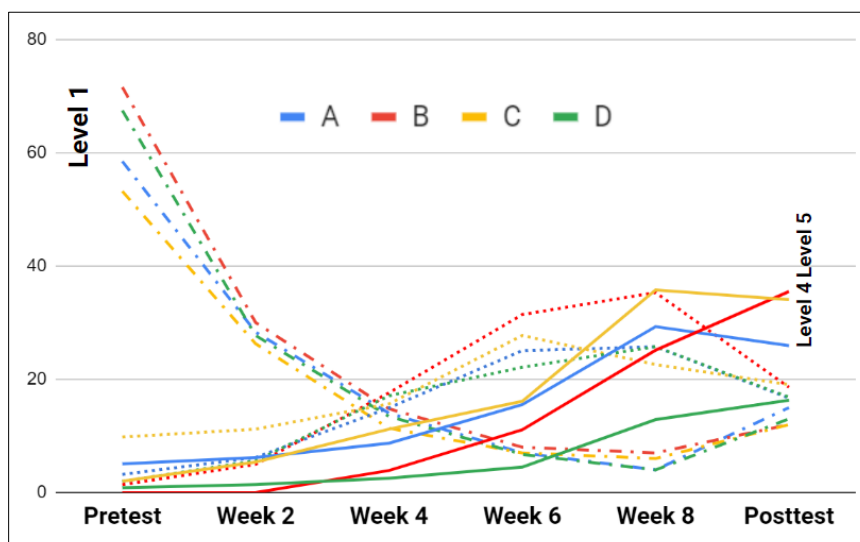
Table 3: Variations in Implementation

Program Implementation	Groups A and C (Group Session)	Groups B and D (Group + Individual Session)
Teachers could not schedule afterschool group remedial sessions due to limited time and/or space	85%	83%
Group remedial sessions were conducted during math classes	85%	66%
Additional group remedial and individual tutoring sessions were carried out at the same time and in the same place (instead of in a separate session and room)	77%	50%
Individual tutoring sessions were conducted during math classes	N/A	66%

Note: The percentage is of the total number of teachers who responded to the endline survey.

Third, in eight weeks, student numeracy outcomes improved significantly across all four intervention groups (Figure 1). The program led to a 50% reduction in the proportion of Level 1 students across all four groups. This implies that the proportion of students at higher levels increased significantly, with a 26% increase in Level 5 (Table 3). When the research team visited a few of the schools during Week 6 of the implementation, schools were having exams, which disrupted normal teaching and learning, including the program implementation. This might explain the drop in learning progression we see towards the end of the intervention.

Figure 1: Percentage of Students in Different Levels during Implementation



Fourth, our quantitative data analysis found significant improvements in student numeracy levels across all four intervention groups. The interventions that complemented group targeting with individual tutoring were on average better than the group targeting intervention alone in reducing the proportion of students at Level 1. However, descriptive statistics indicate two interventions performed the best: Group B, where trained teachers implemented the group and individual targeting sessions, and Group C, where trained school principals supported trained teachers in implementing the group targeting sessions (Table 4). All four groups experienced statistically significant improvement in the average level over the pretest. When we combined the groups based on group targeting only (A and C) and group and individual targeting (B and D), there is no statistically significant difference in the average level of combined groups. Similarly, there is no significant difference between combined groups of teacher training only (A and B) and teacher training with principal support (C and D).

Table 4: Average Level of Students at Pretest and Four Groups at Posttest

Interventions	N	Level Average	P-value*
Pretest	1,545	1.62	–
Group A (Basic only)	366	3.13	0.00
Group B (Basic+ Tutoring)	355	3.47	0.00
Group C (Basic+ Principal)	446	3.48	0.00
Group D (Basic+ Tutoring+ Principal)	356	2.98	0.00
Groups A and C	812	3.32	0.18
Groups B and D	711	3.23	
Groups A and B	721	3.30	0.59
Groups C and D	802	3.26	

* For each intervention group, the average level is compared with the average level of all students at the pretest.

Regression analysis suggests similar findings (Table 5). Using learning level as a continuous variable, the model utilizes an ordinary least squares regression method. Findings suggest that the average learning level of students in Groups B and C is statistically significantly (at 99%) higher by 1.86 points compared to the average level

at the pretest. Results also suggest that girls' average learning level is statistically significantly (at 95%) higher by 0.11 points compared to boys at posttest. The average learning level of Grade 4 students is 0.16 points higher than that of Grade 3 students, statistically significant at 99%.

Table 5: Results of Regression Analysis

	Learning Level
Group A	1.51 (19.33)***
Group B	1.86 (23.22)***
Group C	1.86 (26.30)***
Group D	1.37 (19.37)***
Girls	0.11 (3.78)**
Grade 4	0.16 (2.59)***
Constant	1.48 (40.53)***
<i>R Square</i>	33.94
<i>N</i>	3,068

Notes: $p < 0.1^*$, $p < 0.05^{**}$, $p < 0.01^{***}$; ordinary least squares (OLS) regression method used. The excluded category is the average learning level at pretest. z-score in parentheses. Robust standard error used.

In summary, while the program led to a significant improvement in student numeracy levels overall, we do not see a clear difference between the groups. Interviews with teachers revealed that those in an additional individual targeting intervention found it easier to implement and better for helping students who were struggling with mathematical operations. Teachers highlight that they could exercise more flexibility in providing individual tutoring than in organizing group remedials. For example, a Grade 3 teacher stated: “...I did not have problems with individual tutoring, the parents were supportive if I gave the students individual tutoring. I gave individual tutoring by giving them a quiz through WhatsApp and the next day we would discuss it at school (one to one)...”

Fifth, trained school principals have a positive impact on teacher punctuality in inputting student assessment data by the stipulated deadline, which indicates that the trained principals could better hold teachers accountable than nontrained principals. By Week 8, 64% and 58% of teachers in Groups C and D, respectively, had input student assessment data on time, compared to just 25% and 17% in Groups A and B, respectively. Qualitative interviews with the teachers found that those supported by untrained school principals complained about the fact that the principals lacked information about the intervention and teachers were more likely to contact Inspirasi for any help or clarification.

Table 6: Percentage of Teachers who Submitted Student Assessment Data by the Deadline

	Week 2	Week 4	Week 6	Week 8
A	83	33	25	25
B	58	33	33	17
C	86	57	50	64
D	100	83	58	58

When school principals were involved, students performed better through the group targeting intervention than they did when they received additional individual targeting. As previously mentioned, most schools struggled with fitting in additional group remedial and individual tutoring time due to logistical and scheduling challenges. Observations of check-in meetings between principals and teachers indicated that some principals struggled with supporting teachers in finding schedules for group remedial sessions. Scheduling the additional tutoring sessions for ten students every two weeks requires more complex management skills that some principals may be lacking in. It is possible that teachers in the additional individual targeting group found it discouraging that the school principals could not provide them with the support they needed.

Sixth, in addition to instructional leadership skills, school principals need additional training on management and soft skills. Observations of check-in meetings between school principals and teachers indicated that some of the principals utilized the guiding questions as a checklist, rather than really listening to teachers' challenges and helping them solve problems that teachers raised. A few school principals could not access the online documents utilized for the implementation, which undermined their authority in front of the teachers. For example, one school principal remarked that *"...maybe, I would prefer collecting data manually (on paper) because I can see better with paper than a computer."* Similarly, another principal commented: *"...I think, I prefer to do it manually. Because when I input the data online, I forget where I keep the data, but when I keep the data in a notebook, I find it easier to remember."*

5. CONCLUSIONS AND RECOMMENDATIONS

Across the board and over a short period of time, targeted learning significantly improved students' numeracy levels. After one day of training, teachers exposed to differentiated instruction could implement their newly acquired skills. In general, group targeting complemented by individual targeting was, on average, better than the group targeting intervention alone in reducing the proportion of students at Level 1. However, if we look at the average learning level across groups, Groups B and C performed the best. The participation of trained school principals in the implementation helped in holding teachers more accountable. Group targeting seems to be more effective than when combined with individual targeting in the trained principal interventions.

Based on our experience with this study, we make the following recommendations for future research in this direction. First, most of the schools struggled to add the group and individual targeting after school hours. We recommend that these types of interventions receive endorsement from the government such that the group and individual targeting can be implemented more formally as part of the regular school hours. While this study finds that additional individual tutoring helps improve student outcomes, teachers also struggled in finding the time to do so. Hence, future studies should explore other ways in which tutoring could be provided by nonteachers, including fellow students and volunteers (parents, older siblings, and community members). Second, the program implementation over eight weeks had already improved learning outcomes significantly. Future interventions may want to prolong the timeline to 12 weeks to potentially strengthen the impact on moving students to higher levels, to instill the differentiated teaching practice more strongly, and to encounter potential disruptions due to other priorities schools may have. Third, future training content for teachers should include overcoming biases in students' ability to progress. Training for school principals should cover managerial aspects (such as scheduling

time tables) and soft skills (such as conducting reflection meetings and problem solving), in addition to using data to inform support for teachers.

Training content should be scaffolded for both teachers and school principals in such a way that they get to practice what they learn first, have a debriefing session with their peers, and then receive additional and more challenging content later. For example, group targeting could be introduced in the first training session for teachers, followed by a facilitated debriefing session after they have implemented the session once to share confusions and challenges, and get clarifications and solutions. In the second training session, they can then identify students who would need individual targeting. Teachers should be able to better implement the group targeting classes and, in addition, add individual targeting for those needing one-on-one guidance. The training sessions for school principals should focus not only on providing them with the instructional leadership skills needed to lead these interventions, but also on relevant management skills, such as listening and responding to problems raised by teachers, and reconfiguring timetables to fit in the group and individual targeting sessions. Given the challenges that some principals and teachers faced with digital literacy, adding an opt-in, hands-on session may help them improve their skills.

REFERENCES

- Afkar, R., and N. Yarrow. 2021. Rewrite the Future: How Indonesia's Education System Can Overcome the Losses from the COVID-19 Pandemic and Raise Learning Outcomes for All. The World Bank. <https://openknowledge.worldbank.org/entities/publication/1251eca8-7505-519a-86ed-e6dadbab34d3>.
- Angrist, N., et al. (2023). Building Resilient Education Systems: Evidence from Large-Scale Randomized Trials in Five Countries (No. w31208). National Bureau of Economic Research. <https://www.nber.org/papers/w31208>.
- Angrist, N., P. Bergman, and M. Matsheng. 2022. Experimental Evidence on Learning Using Low-tech When School Is Out. *Nature Human Behaviour* 6(7): 941–950. <https://doi.org/10.1038/s41562-022-01381-z>.
- Angrist, N., P. Bergman, C. Brewster, and M. Matsheng. 2020. Stemming Learning Loss During the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention in Botswana. Social Science Research Network. <https://doi.org/10.2139/ssrn.3663098>.
- Azevedo, J. P., A. Hasan, D. Goldemberg, S. A. Iqbal, and K. Geven. 2020. Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes: A Set of Global Estimates (Policy Research Working Paper 9284). The World Bank. <https://documents1.worldbank.org/curated/en/329961592483383689/pdf/Simulating-the-Potential-Impacts-of-COVID-19-School-Closures-on-Schooling-and-Learning-Outcomes-A-Set-of-Global-Estimates.pdf>.
- Banerjee, A. et al. 2016. Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of “Teaching at the Right Level” in India. National Bureau of Economic Research, NBER Working Paper 22746. <https://doi.org/10.3386/w22746>.
- . 2017. From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. *Journal of Economic Perspectives* 31(4): 73–102. <https://doi.org/10.1257/jep.31.4.73>.
- Banerjee, A., S. Cole, E. Duflo, and L. L. Linden. 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics* 122(3): 1235–1264. <https://doi.org/10.1162/qjec.122.3.1235>.
- Beatty, A., E. Berkhout, L. Bima, M. Pradhan, and D. Suryadarma. 2021. Schooling Progress, Learning Reversal: Indonesia's Learning Profiles Between 2000 and 2014. *International Journal of Educational Development* 85: 102436. <https://doi.org/10.1016/j.ijedudev.2021.102436>.
- Beg, S., A. Fitzpatrick, and A. M. Lucas. 2021. Improving Public Sector Service Delivery: The Importance of Management [Conference Paper]. 6th IZA Workshop: The Economics of Education, United States of America. https://conference.iza.org/conference_files/edu_2021/beg_s31721.pdf.
- Cilliers, J., E. Dunford, and J. Habyarimana. 2021. What Do Local Government Education Managers Do to Boost Learning Outcomes? (RISE Working Paper 21/064). RISE. https://doi.org/10.35489/BSG-RISEWP_2021/064.

- Crawford, L., D. K. Evans, S. Hares, and J. Sandefur. 2021. Teaching and Testing by Phone in a Pandemic (Working Paper 591). Center for Global Development. <https://www.cgdev.org/sites/default/files/teaching-and-testing-phone-pandemic.pdf>.
- De Barros, R. P., M. De Carvalho, S. Franco, B. Garcia, R. Henriques, and L. Machado, (2022). Assessment of the Impact of the Jovem de Futuro Program on Learning. The World Bank. <https://documents1.worldbank.org/curated/es/825101561723584640/pdf/Assessment-of-the-Impact-of-the-Jovem-de-Futuro-Program-on-Learning.pdf>.
- Duflo, A., J. Kiessel, and A. M. Lucas. 2020. Experimental Evidence on Alternative Policies to Increase Learning at Scale. Social Science Research Network. <https://doi.org/10.3386/w27298>.
- Duflo, E., P. Dupas, and M. Kremer. 2008. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. National Bureau of Economic Research, NBER Working Paper No. 14475. https://www.nber.org/system/files/working_papers/w14475/w14475.pdf.
- . 2011. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review* 101(5): 1739–1774. <https://doi.org/10.1257/aer.101.5.1739>.
- Grissom, J. A., A. J. Egalite, and C. Lindsay. 2021. How Principals Affect Students and Schools: A Systematic Synthesis of Two Decades of Research. The Wallace Foundation. <https://www.wallacefoundation.org/knowledge-center/pages/how-principals-affect-students-and-schools-a-systematic-synthesis-of-two-decades-of-research.aspx>.
- Merdeka B. | BKHM | KEMDIKBUD. (n.d.). http://merdekabelajar.kemdikbud.go.id/episode_15/web.
- Moscoviz, L., and D. K. Evans. 2022. Learning Loss and Student Dropouts during the COVID-19 Pandemic: A Review of the Evidence Two Years after Schools Shut Down (Working Paper 609). Center for Global Development. <https://www.cgdev.org/sites/default/files/learning-loss-and-student-dropouts-during-covid-19-pandemic-review-evidence-two-years.pdf>.
- Muralidharan, K., and A. Singh. 2020. Improving Public Sector Management at Scale? Experimental Evidence on School Governance India. National Bureau of Economic Research, NBER Working Paper No. 28129. https://www.nber.org/system/files/working_papers/w28129/w28129.pdf.
- Muralidharan, K., A. Singh, and A. Ganimian. 2017. Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. National Bureau of Economic Research, NBER Working Paper No. 22923. https://www.nber.org/system/files/working_papers/w22923/w22923.pdf.
- PISA 2018 Results (Volume I). 2019. Programme for International Student Assessment. <https://doi.org/10.1787/5f07c754-en>.
- RISEProgramme. 2022, June 24. RISE Annual Conference 2022 – Day 2 [Video]. YouTube. <https://www.youtube.com/watch?v=fr45M9Pa-QY>.

- Romero, M., J. Bedoya, M. Yanez-Pagans, M. Silveyra, and R. De Hoyos. 2021. School Management, Grants, and Test Scores: Experimental Evidence from Mexico (Policy Research Working Paper 9535). The World Bank. <https://openknowledge.worldbank.org/server/api/core/bitstreams/70f2df5a-cb89-5726-bc56-f9f4e36cf7d9/content>.
- Scur, D., R. Sadun, J. Van Reenen, R. Lemos, and N. Bloom. 2021. World Management Survey at 18: Lessons and the Way Forward (IZA DP No. 14146). IZA – Institute of Labor Economics. <https://docs.iza.org/dp14146.pdf>.
- Singh, A., Romero, M., & Muralidharan, K. (2022). COVID-19 Learning loss and recovery: Panel data evidence from India (No. w30552). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w30552/w30552.pdf.
- UNESCO. 2022, June 29. Education: From Disruption to Recovery. <https://webarchive.unesco.org/web/20220629024039/https://en.unesco.org/covid19/educationresponse/>.
- UNICEF. 2021. Towards A Child-focused COVID-19 Response and Recovery: A Call to Action. <https://www.unicef.org/indonesia/media/10666/file/Towards%20a%20child-focused%20COVID-19%20response%20and%20recovery.pdf>.

ANNEXURES

Table 1: Proportion of Students at Levels 1–5 through Five Assessment Cycles

	Pretest	Week 2	Week 4	Week 6	Week 8	Posttest
Level 1						
A	58.51	27.96	13.59	7.36	4.11	15.03
B	71.63	29.83	14.33	8.38	6.59	12.39
C	53.23	26.4	11.41	7.38	5.82	11.88
D	67.51	27.73	13.45	6.72	3.92	13.48
Level 2						
A	24.73	41.94	34.24	24.25	16.44	25.14
B	20.39	39.78	24.78	17.07	12.87	18.03
C	22.05	40.04	32.44	22.15	10.07	15.7
D	22.97	38.94	36.13	28.29	21.29	23.88
Level 3						
A	8.51	17.74	28.53	27.79	24.38	17.21
B	6.61	25.41	39.4	32.04	20.06	15.49
C	12.92	17.00	29.31	26.62	25.73	19.28
D	6.72	26.33	30.81	38.38	36.13	29.49
Level 4						
A	3.19	6.18	14.95	25.07	25.75	16.67
B	1.38	4.97	17.61	31.44	35.33	18.59
C	9.8	11.19	15.66	27.74	22.6	19.06
D	1.96	5.6	17.09	22.13	25.77	16.85
Level 5						
A	5.06	6.18	8.7	15.53	29.31	25.96
B	0	0	3.88	11.08	25.15	35.5
C	2.00	5.37	11.19	16.11	35.79	34.08
D	0.84	1.4	2.52	4.48	12.88	16.3
Total Students	1,545	1,538	1,507	1,505	1,503	1,523

Table 2: Percentage of Boys and Girls at Levels 1–5 at Pretest and Posttest

		Level 1	Level 2	Level 3	Level 4	Level 5
Boys	Pretest	64.22	21.73	8.67	3.91	1.46
	Posttest	14.99	19.83	20.57	17.6	27.01
	Difference	-49.23	-1.9	11.9	13.69	25.55
Girls	Pretest	59.78	23.42	9.23	4.96	2.61
	Posttest	11.03	21.09	19.97	18.16	29.75
	Difference	-48.75	-2.33	10.74	13.2	27.14

**Table 5: Percentage of Grade 3 and 4 Students at Levels 1–5
at Pretest and Posttest**

		Level 1	Level 2	Level 3	Level 4	Level 5
Grade 3	Pretest	65.78	21.09	8.09	3.58	1.46
	Posttest	15.17	19.73	21.07	18.39	2.53
	Difference	-50.61	-1.36	12.98	14.81	1.07
Grade 4	Pretest	58.66	23.89	9.73	5.18	25.64
	Posttest	11.18	21.08	19.54	17.35	30.85
	Difference	-47.48	-2.81	9.81	12.17	5.21