Final Report

Guru BAIK: building teachers' capacity in West Nusa Tenggara, Indonesia

April 2019

INOVASI – Innovation for Indonesia's School Children

Ratu Plaza Office Tower 19th Floor,

Jl. Jend. Sudirman Kav 9, Jakarta Pusat, 10270

Indonesia

Tel : (+6221) 720 6616 ext. 304

Fax : (+6221) 720 6616

http://www.inovasi.or.id

April 2019

Cover photo courtesy of Palladium

The Governments of Australia and Indonesia are partnering through the Innovation for Indonesia's School Children (INOVASI) program. INOVASI seeks to understand how to improve student learning outcomes in literacy and numeracy in diverse schools and districts across Indonesia. The first phase of the Program (AUD49 million) began in January 2016 and will continue until December 2019. Working with Indonesia's Ministry of Education and Culture, INOVASI has formed partnerships with 17 districts in: West Nusa Tenggara; Sumba Island, East Nusa Tenggara; North Kalimantan; and East Java.

INOVASI is an Australia–Indonesia Government Partnership – Managed by Palladium.

Final Report

Guru BAIK: building teachers' capacity in West Nusa Tenggara, Indonesia

April 2019

CONTENTS

LIST OF TABLES AND FIGURES

LIST OF ACRONYMS

DAPODIK	basic edication data (*Data Pokok Pendidikan*)
GB	Guru BAIK
INAP	Indonesian National Assessment Program (INAP)
INOVASI	Innovation for Indonesia's School Children
IRT	item response theory
KKG	teachers' working groups (*kelompok kerja guru*)
LPMP	Educational Quality Assurance Council (LPMP)
MERL	monitoring, evaluation, research and learning
MoEC	Ministry of Education and Culture
NTB	West Nusa Tenggara (*Nusa Tenggara Barat*)
PAUD	early childhood education (*pendidikan anak usia dini*)
PDIA	problem-driven iterative adaptation approach
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PNS	civil servant
REDI	Regional Economic Development Institute
RPP	lesson plan (*rencana pelaksanaan pembelajaran*)
SD	Primary school (sekolah dasar)
SDIT	integrated Islamic private school (sekolah dasar Islam terpadu)
SDN	state schools (sekolah dasar negeri)
SDS	private school (sekolah dasar swasta)
SIPPI	Indonesian education and learning innovation survey
SMP	junior secondary school (*sekolah menengah pertama*)
SR	Richter scale (SR)
TIMSS	Trends in International Mathematics and Science Study (TIMSS)
TK	kindergarten

EXECUTIVE SUMMARY

Indonesia has performed well in increasing access to education, particularly at the primary school level. Enrolment has reached universal levels, as illustrated by the national indicators of both gross and net enrolment rates. However, the increased enrolment rates are not parallel with increased student learning outcomes. Results from the Programme for International Student Assessment (PISA) suggest that Indonesian students' achievements are below those of students in other developing countries in the same region. The results from the Indonesian National Assessment Program (INAP) confirm that the students' learning outcomes are still below expectation. Thus the government is focusing on improving the quality of learning and teaching to improve students' learning outcomes.

Innovation for Indonesia's School Children (INOVASI) is a joint education program, funded by the Australian government in partnership with the Indonesian Ministry of Education and Culture (MoEC), to support the Government of Indonesia in improving student performance, particularly in literacy and numeracy in the early grades. INOVASI works to understand and tackle the learning challenges related to literacy and numeracy in classrooms and schools in Indonesia. The first phase of INOVASI runs from 2016 to 2019.

INOVASI operates through pilots that test what interventions work in certain contexts to improve learning outcomes. A variety of pilots have been implemented across 17 partner districts in four provinces. Guru BAIK was the first pilot and was implemented in North Lombok and Sumbawa districts from January 2017 to May 2017. Guru BAIK aims to build the capacity of teachers, equipping them with the knowledge and skills to integrate action research principles into their teaching and problem-solving methods, in order to tackle immediate issues and challenges with regard to literacy and numeracy in their classrooms. The pilot, based on the principles of classroom action research, consists of a series of workshops and includes a multi-stage, problem-driven, cyclical process of identifying problems, planning action to address the problems, taking action and collecting data, analysing the results of the action, reflecting and re-planning the action.

A total of 50 schools participated in Guru BAIK from across North Lombok and Sumbawa, with 25 schools from each district. All schools were public schools, except for one Islamic school. All schools were also under the care of MoEC. At the time of the pilot, INOVASI worked exclusively with MoEC, before later collaborating with the Ministry of Religious Affairs (MoRA).

Prior to implementing the pilots, baseline studies were conducted for the Guru BAIK and Gema Literasi pilots at the same time. Gema Literasi is a literacy project implemented by Save the Children as INOVASI's partner. It focuses on improving students' literacy by enhancing teachers' capacity in literacy instruction and increasing community support. The beneficiaries of Guru BAIK are teachers whereas Gema Literasi also targets parents. The endline studies for both pilots were administered in August 2018. However, due to the earthquakes that hit West Nusa Tenggara (NTB) in August 2018, the endline study had to be cancelled in North Lombok, leaving the focus on Sumbawa. Sumbawa was later also hit by earthquakes so the endline study could not be completed in all the Guru BAIK and Gema Literasi partner schools. Consequently, this report is based on the data collected from some of the schools in Sumbawa. Meanwhile, another study was conducted that bridged the gap of knowledge on the Guru BAIK and Gema Literasi programs in North Lombok and at the same time responded to the current local needs. This was a study on disaster and how the teachers responded to the situation. Data collection took place from December 2018 to January 2019.

This report presents the main findings from the endline survey of the Guru BAIK pilot and includes some key and comparable findings from the Gema Literasi endline study and the disaster study although separate reports have been developed for each study. In addition, the Guru BAIK monitoring data and midline study are incorporated in this report. The findings illustrate the changes that took place after the pilot was completed. These changes are identified by comparing the baseline and endline results across Guru BAIK (as the treatment group), Gema Literasi (as another treatment group) and the control schools (with no treatment).

Overall, the endline study revealed that the Guru BAIK schools performed better than the Gema Literasi and control schools although there were variations in some of the variables.[1] Students' literacy and numeracy scores in the Guru BAIK schools were 1.95 and 2.92 points higher respectively than the control group although the effect size was small (0.21 and 0.25, respectively). For Gema Literasi schools, the literacy scores were also higher than for the control group at 0.578 with an effect size of 0.06. Thus, statistically, the net effect was smaller than for Guru BAIK and there was no significant difference between Gema Literasi and the control group. However, using Gema Literasi-specific indicators, such as the number of students who met the minimum passing grades, Gema Literasi schools performed better than the control group.

Guru BAIK was found to have a greater effect on boys compared to girls, especially for numeracy (with effect sizes of 0.25 and 0.17 respectively for literacy; and 0.33 and 0.17 respectively for numeracy). When the results were disaggregated by grades, the effect was the highest among grade one students both for literacy and numeracy, and gradually decreased up to grade three. For grade four students, the effect was negative in the case of literacy and positive in the case of numeracy. Similarly, the effect of Gema Literasi was greater for boys than for girls although the effect size was insignificant.

Further investigation is required to understand why Guru BAIK generated stronger results for boys than girls. Considering the current data and evidence, two assumptions are proposed. First, baseline data showed that girls learning outcomes were better than boys. Thus, the starting point for girls was higher than boys. Considering the evidence that the Guru BAIK pilot impact decreased at a higher level of learning, this may explain why net change for girls was less than boys. It is assumed that Guru BAIK worked better for boys as boys were at a lower learning level initially compared to girls. Secondly, Guru BAIK drives teachers to understand teaching and learning problems in their classrooms, for individual students. Teachers were trained to identify issues, work with data of, for example, formative assessment, and do reflection. Then, teachers could better identify needs and issues in the classroom. It is assumed that teachers ended up working more with children who had greater learning needs; based on baseline findings, girls already outperformed boys and thus may not have needed greater initial assistance. Having said that, additional data collection would be needed to test these assumptions.

While students' scores clearly improved for both subjects among students in the Guru BAIK group, comparing other variables related to students' attitudes resulted in greater variations. For instance, the net change difference[2] in students' enthusiasm for school after the Guru BAIK intervention was negative and the proportion difference in the endline survey was statistically significant. There was an increase in the proportion of students who chose Bahasa Indonesia as their favourite subject in the Guru BAIK group while this proportion decreased in the control group, leading to a positive net change difference for Guru BAIK. Meanwhile, in both groups, students' attitudes towards mathematics tended to have declined in the endline survey resulting in a negative net change difference for Guru BAIK.

Indicators reflecting students' attitudes from the perspectives of parents and principals also tended to show a net negative change. However parents were happy with the indicators of their children's development, including: literacy and numeracy skills, having friends, identifying role models; and better behaviour. The findings suggest that more parents were satisfied in all of these aspects.

The focus for teachers was on improved teaching practices, knowledge, attitudes and skills. Their teaching practices were assessed on the teachers' use of active learning and learning media, the degree of student engagement, the types of assignments they gave students and how often teachers were absent. The survey found that teachers increasingly used active learning techniques in both the Guru BAIK and the control groups, with a net positive effect of 1 per cent for the Guru BAIK schools.

[1] Student scores were analysed using the item response theory (IRT) approach and the two parameter logistic model considering discrimination power and item difficulties.

[2] Net change difference of an indicator is calculated by substracting the change of the indicator's value from the baseline to the endline surveys among the control group from the change of the same indicator's value from the baseline to the endline surveys among Guru BAIK group. Mathematically, this can be written as $\Delta y_{net} = \Delta y_{treated} - \Delta y_{control}$, where y denotes the respective indicator.

Also, the proportion of classes where students were engaged increased significantly (by 6 per cent) in the Guru BAIK group while this proportion decreased by more than 14 per cent in the control group resulting in a high net effect of up to 21 per cent. This positive net effect in students' engagement might relate to the teachers increasingly using active instruction approaches.

Teachers' use of learning media is equally important but the practice is still limited. Only around 14 per cent and 3 per cent of the observed classes used learning media in Guru BAIK and the control groups respectively. While there was still a positive trend in using learning media among the Guru BAIK schools, in the control schools this had declined by 3 per cent, leading to a positive net change of 13 per cent for the Guru BAIK group.

The three types of assignments teachers gave students that the study focused on were: (1) reading from other sources; (2) putting what was learned into practice; and (3) expressing opinions about subjects. The proportion of teachers who asked students to perform these three types of assignment every day was assessed. The results showed positive net changes for the three types of assignment and indicated the positive effects of Guru BAIK interventions by 26 per cent, 33 per cent and 22 per cent respectively for each type of assignment. In addition, a smaller proportion of teachers were absent among the Guru BAIK group compared to the control group in the endline survey.

Teachers' participation in teachers' working groups (KKG) and in various training programs were important indicators of their development. Overall, the Guru BAIK interventions had a positive effect on teachers' development. There was an increasing trend of teachers participating in teachers' working groups from the baseline to the endline surveys and a significantly higher proportion of these teachers were from the Guru BAIK group in the endline survey. These findings were possibly related to teachers' overall satisfaction levels with the teachers' working groups, as the results showed a positive net change difference for the treatment group.

Another contrasting finding between the Guru BAIK and control groups was the proportion of teachers being assigned for training. While the percentage increased by around 13 per cent in the Guru BAIK group, it decreased by 3 per cent in the control group, resulting in a positive net change difference of around 16 per cent. A similar trend was found with regard to teachers' satisfaction with their workloads, with a 16 per cent increase for the Guru BAIK group and small decrease of about 2 per cent for the control group.

Teachers' knowledge was directly reflected in their subject-knowledge test scores. There was a notable net change difference in teachers' average literacy score between the Guru BAIK and the control groups, as the scores increased for the Guru BAIK teachers but decreased in the control group, becoming statistically significant in the endline survey. On the other hand, the net change difference in teachers' numeracy scores between the Guru BAIK and the control groups was negative. This counterintuitive finding with regard to teachers' numeracy scores calls for further inquiry.

Apart from the subject knowledge component, teachers' shift in mindset was captured by two proxies: whether they used data to identify school problems and whether they thought that their teaching competencies should be improved. With regard to the first proxy, a higher proportion of teachers used data to identify school problems in both the treated and the control groups in the endline survey but the net change was positive for the Guru BAIK group. A similar trend was observed for the second proxy where more teachers thought their teaching competencies should be improved, with a higher proportion and net change in the Guru BAIK group than in the control group. These two indicators suggest that teachers observed and learned from what they delivered in the classroom and built their knowledge through the data they collected. Secondly, they indicate that teachers realise they are part of the solution, are aware of the need to improve and see this as an area they can deal with or take the initiative.

Stakeholders' perspectives on teachers' knowledge, attitude and skills were not always compatible with the observed improvements. Among parents, there were decreasing trends in their perceptions of the quality of teachers. Overall, there were decreasing trends in school principals' satisfaction with several aspects for both

groups, although some net changes were in favour of Guru BAIK. For instance, school principals were less satisfied with teachers' understanding of the aim of the curriculum in both groups in the endline survey although the decrease in the control group was higher. Secondly, there were negative net change differences for the Guru BAIK group with regard to the satisfaction with teachers' knowledge and pedagogical competence, as well as with students' learning outcomes.

As mentioned, there was a separate study for Guru BAIK and Gema Literasi schools in North Lombok after the earthquakes. Overall, there were distinctive results between the Guru BAIK and Gema Literasi teachers on one side and the control group on the other in performing their teaching roles during the recovery stage. Firstly, both Guru BAIK and Gema Literasi teachers applied the practice of developing adjusted lesson plans while the study could not identify similar practices within the control group. However, one difference was noted between Guru BAIK and Gema Literasi teachers in developing lesson plans: Guru BAIK teachers did a situation analysis to identify their students' problems and put efforts into getting children to recall the lessons they had learned before the disaster. This approach resulted in simpler, more relevant and adaptive lesson plans compared to the more rigid conventional plans. Gema Literasi teachers emphasised literacy in their lesson plans, covering activities, such as reading fairy tales and presenting students' work.

In terms of teaching practices, Guru BAIK teachers used the material or media available in their surroundings and paid attention to the psychological state of their students. Similarly, Gema Literasi teachers used materials available from nearby schools. The materials and facilities they were used to had been largely destroyed by the earthquakes and the lack of supporting infrastructure, like reading corners and reading camps, hindered their usual approaches. Nevertheless they focused on boosting literacy and maintained the 15-minute reading practice habit. The students were encouraged to produce and present their work in their temporary classrooms. Such practices were not found in the control schools where most of the teachers use conventional teaching methods and put students under pressure to study the lesson quickly to meet the targets and not lag behind.

These differences were also found in the teachers' shift in mindset. Compared to teachers in the control schools, Guru BAIK and Gema Literasi teachers tended to have the enthusiasm and determination to teach, even in an emergency situation. They changed their teaching approaches to be more creative and focused on finding solutions to the problems they faced due to the earthquakes, rather than waiting for instructions from the government or related agencies. In contrast, control schools tended to await instruction. Guru BAIK teachers, in addition to having a more creative mindset, also tried a variety of learning approaches that were suitable for their students.

To conclude, clear differences emerged between Guru BAIK and Gema Literasi, in one side, and the control groups, in the other side, in various aspects of teaching and learning, during both normal and emergency situations. While Guru BAIK and Gema Literasi teachers were similar in certain aspects, Guru BAIK generated higher performance particularly in student learning outcomes. Some findings merit further elaboration to investigate 'why' and 'what particular contexts' contribute to such results.

1. INTRODUCTION

1.1 BACKGROUND

Innovation for Indonesia's School Children (INOVASI) is an education program funded by the Australian Government in partnership with the Indonesian Ministry of Education and Culture (MoEC). The first phase of the program runs from 2016 to 2019. INOVASI works to understand and tackle learning challenges in classrooms and schools, and particularly those relating to literacy and numeracy. The program's three focus areas are:

1 strengthening the quality of teaching and learning in the classroom;

2 improving the support provided to teachers; and

3 enabling all children in the classroom to reach their potential in learning.

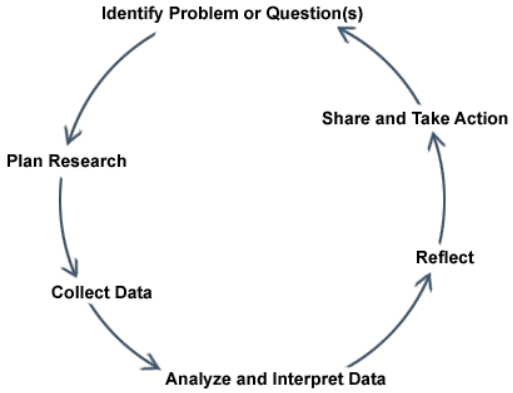
INOVASI implemented its first pilot, Guru BAIK, in West Nusa Tenggara (NTB) province. *Guru* means 'teacher' in Bahasa and BAIK stands for *'belajar, aspiratif, inklusif dan kontekstual'* which means 'aspirational, inclusive and contextual learning'. The pilot aimed to build the capacity of teachers, equipping them with the knowledge and skills to integrate action research principles into their teaching and problem-solving methods, in order to tackle immediate issues and challenges with regard to literacy and numeracy in their classrooms.

This report presents the results from the endline survey of Guru BAIK in West Nusa Tenggara. The report provides comprehensive information about the situation of students, teachers and schools after the pilot was implemented. This first chapter describes Guru BAIK and the strategies used to measure change indications due to the intervention. Chapter two presents the data collection process and reports on implementing the survey. Chapter three provides the findings from the endline study on students, teachers, school principals, school supervisors and parents, respectively. It also compares the indicators before and after the pilot was implemented, as well as the performance of the treated and controlled groups. Overall, the findings indicate some improvements in the situation of students, teachers, school principals, school supervisors and parents that can potentially be associated with the Guru BAIK pilot intervention. Analyses of the correlating factors of students' literacy and numeracy performances using the regression technique are also described. The final chapter sums up all the findings.

1.2 PILOT PLAN

Guru BAIK (GB) focuses on improving the quality of classroom learning by supporting teachers in developing their competencies to solve classroom learning problems. Guru BAIK uses problem-driven iterative adaptation (PDIA), the same underlying principle as the overall INOVASI program. This gives teachers the knowledge and skills to integrate action research principles as a contextual problem-driven methodology to tackle immediate issues and challenges with regard to literacy and numeracy in their classrooms. Figure 1.1 shows the pilot activity cycle for Guru BAIK.

Figure 1. 1. Cycle of Guru BAIK pilot activities



Implementing Guru BAIK involved a series of four workshops with connected, guided and mentored follow-on activities. At the end, a workshop was conducted to disseminate the innovations or good practices generated during the pilot.

1. **Workshop one and follow-on activities (identify problems or questions):** During the first workshop, teachers are given support in identifying a problem, challenge or research question they have with literacy and numeracy in their own classrooms. After the workshop, teachers review and confirm the research questions.
2. **Workshop two and follow-on activities (plan research):** In the second workshop, teachers start planning out how they could solve the problems or challenges or answer their research questions. They are given support in conducting a literature review, discussing possible solutions with peers and identifying existing promising practices. Following the workshops, teachers continue to develop their action plans.
3. **Workshop three and follow-on activities (plan research and collect data):** This workshop helps teachers develop a methodology (and accompanying instruments) for collecting and organising the data they need to measure whether and to what extent their action plans have been successful. After this workshop, teachers carry out their research according to their plans and collect the necessary data.
4. **Workshop four and follow-on activities (analyse and interpret data and reflect):** The final workshop helps teachers analyse and interpret the data they have collected to see whether and to what extent their actions have been successful. As a follow up to this workshop, all participants are expected to repeat the cycle or start again with new questions or problems found in their research.
5. **Dissemination (share findings and take action):** Participating teachers are asked to document (in their own way) and share their findings and experience with teachers in their own schools, in other schools in their clusters and finally across the districts.

Following the Guru BAIK pilot and upon request, INOVASI provides technical assistance to local districts to expand the program to teachers in other schools.

Evidence suggests that the action research process works best through collaboration and cooperation so the Guru BAIK pilot was implemented by 50 research teams in 50 primary schools across two districts (North Lombok and Sumbawa) in West Nusa Tenggara. Research teams comprised three members as follows:

- 1 lead teacher to conduct the research in the classroom;
- 1 teacher from the same school to act as a 'critical friend';
- 1 academic from a local higher education institute to provide oversight on the research methods.

A group of national facilitators with experience in conducting action research in the classroom and a group of local facilitators from the district education office, local university or the Educational Quality Assurance Council (LPMP) mentored and supported each research team.

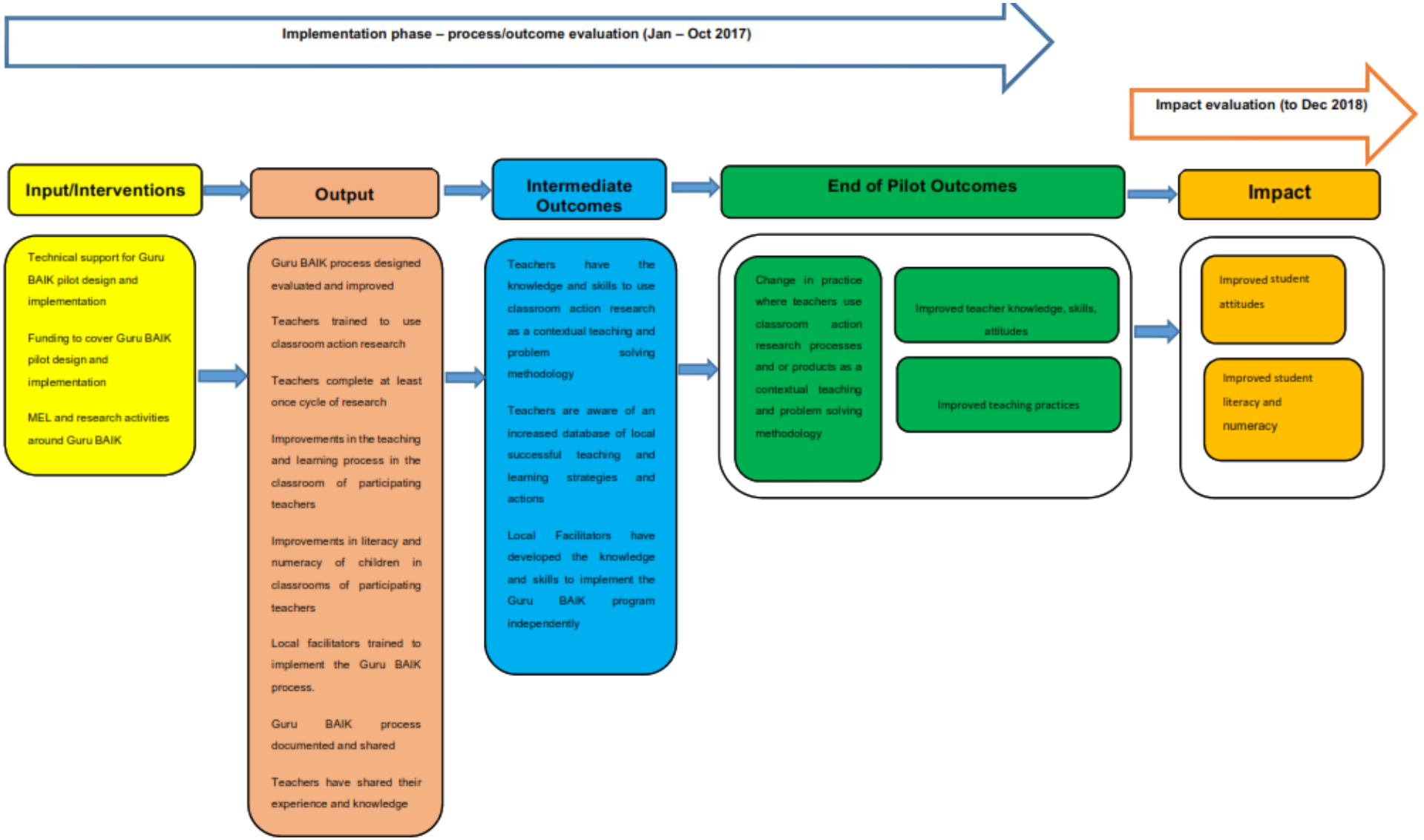
INOVASI and the national facilitators mentor the local facilitators and train them to implement the Guru BAIK pilot. The aim is for them to become core resources in their districts so they can continue to roll out the Guru BAIK program to other teachers and schools.

At the end of the Guru BAIK pilot, INOVASI expected the following results:

1. Participating teachers have the capacity to use action research as a contextual teaching and problem solving methodology to improve students' learning in literacy and numeracy in their classrooms;
2. The quality of teaching and learning has improved in the classrooms of participating teachers;
3. Students' learning outcomes in literacy and numeracy have improved in the classrooms of participating teachers as a result of the actions taken;
4. A collection of locally-relevant promising practices has been developed and can be shared;
5. A core group of local facilitators has been developed that has the capacity to sustain and scale out the Guru BAIK pilot activity;
6. A core group of teachers is committed to continuously using the action research methodology to address any new problems and challenges with literacy and numeracy as they emerge.

Guru BAIK follows a logical theory of change and figure 1.2 depicts how the pilot program is designed to work. In the long-term, INOVASI expects literacy and numeracy levels in basic education to improve in the areas where the program has worked. For this to happen, education stakeholders (categorised as policymakers, intermediaries and practitioners) need to be able to successfully use context-relevant approaches to generate best-fit solutions to problems when working to improve the quality of education. Classroom action research is always relevant to the participants and the context because the focus of every research project is determined by the teachers as the main researchers, practitioners and the primary consumers of the findings. The design for Guru BAIK is underpinned by a simple idea that by designing an action research process and guiding teachers through it stage by stage, teachers will develop the skills to continue to use the process independently. These teachers will understand that they control how they work in the classroom and that finding simple, manageable actions to improve their classroom practices can ultimately improve student learning outcomes. This is an empowering experience for the teachers and when they have convincing evidence that their work and efforts have made a difference, they will continue to use the process.

Figure 1. 2: Theory of change, Guru BAIK



PS: The timeframe for evaluation has been adjusted, from December 2018 to August 2018, to get the findings earlier. Secondly, the term impact has been changed to **effect** of the pilot as the evaluation design is not applying RCT; in addition to this, the pilot schools were not randomly selected.

Guru BAIK is central to achieving INOVASI's end-of-program outcome as it is specifically designed to support practitioners in developing the skills and understanding to independently use a context-driven, problem-solving and iterative process to improve literacy and numeracy. Guru BAIK contributes to achieving the outputs and therefore the intermediate outcomes through several mechanisms:

- The action research that teachers conduct provides a collection of potentially promising practices in improving literacy and numeracy, as well as evidence on the extent to which they work in a specific classroom context.
- By engaging in the Guru BAIK pilot and conducting the guided research project, participants gain practical, hands-on experience in using a context-specific, problem-driven approach.
- Testing and evaluating the Guru BAIK pilot enables the stakeholders to continue to reflect and improve on the methodology used and to document and share how it works.
- Asking teachers to work in groups and share their experiences and findings with other teachers in their schools and clusters using formal and non-formal channels facilitates knowledge sharing among them.

1.3 SCHOOL SELECTION

The endline survey is the continuation of the Indonesian education and learning innovation survey (SIPPI)–2016 West Nusa Tenggara baseline study that took place in two districts in the province, North Lombok and Sumbawa. The study planned to collect data from 150 partner schools, made up of 149 state schools (SDN) and one private school (SDS) at the primary level under the Ministry of Education and Culture (MoEC). The one private school is classified as an integrated Islamic private school (SDIT). According to the research design, out of these 150 schools,[3] 50 were control schools, 50 were intervention schools (Guru BAIK schools) and the rest were schools that participated in the Gema Literasi pilot run by Save the Children. These schools were evenly distributed in North Lombok and Sumbawa.

Due to the earthquakes that severely hit some areas in West Nusa Tenggara in August 2018, data collection for the SIPPI–2018 West Nusa Tenggara endline in North Lombok was put on hold and finally cancelled. In addition, not all schools in Sumbawa were surveyed due to the earthquakes. Consequently the endline study focused on collecting data from 49 of the targeted schools in Sumbawa. Further details of data collection management are explained under section 2.2 on implementing the survey.

2. METHODOLOGY

This chapter elaborates on the protocol used to prepare for the Guru BAIK pilot's endline survey and explains some of the technical issues that arose in implementing the survey. As explained, the pilot has implications for multiple actors and stakeholders, including teachers, school principals, school supervisors, parents, communities and students. Similar to the baseline survey process, specific instruments were used to collect data from each stakeholder group and this chapter describes the content and administration of these instruments in the field.

[3] The schools were chosen with the help of district officials through two-hour workshops held in each district. The INOVASI team first created an index of school readiness in the two districts based on administrative data gathered at the school level. This data, provided by the Ministry of Education and Culture is known as DAPODIK (Data Pokok Pendidikan – core education data). The variables used were: internet access; teacher–student ratios; proportion of civil servant teachers; proportion of good classrooms; availability of a library; water access; whether the school was accredited by the ministry; and the availability of a special needs teacher. The index was then used to stratify the schools into bottom 20 per cent, middle 60 per cent and top 20 per cent. The list of the middle 60 per cent schools was then shared with the district officials and the schools were chosen from this list.

2.1 SURVEY INSTRUMENT

The survey used the SIPPI questionnaires that are the primary source of information on a given respondent. SIPPI has separate questionnaires for students, parents, teachers, school principals, school supervisors and for school and classroom observations. This dataset allows us to control potentially confounding factors and analyse different effects (heterogeneity) across sub-groups. Detailed descriptions of all the survey instruments used in this study are shown in table 2.1.

During the baseline, the survey was done for Grade 1-5, while in the endline study, the survey was administered for Grade 3-6, to continuously assessed the same students. Please note that the survey was conducted after two years of academic calendars. The teacher assessed in the endline were the same one with the respondents during the baseline survey. Number of teachers per school in endline survey was smaller than the baseline as not all of the original respondents were still available or still work in the same schools due to various reasons (for examples due to teacher rotation, pension, and earthquake).

Table 2. 1: Survey instruments

Instruments	Target	Note	Average completion time
Principal survey	1 principal per school		45 minutes
Teacher survey	≥ 2 teachers per school	≥ 4 for baseline study	45 minutes
Teacher test	≥ 2 teachers per school	Mathematics and Bahasa Indonesia, each ≥ 4 for baseline study	90 minutes (both)
Teachers self-administered questionnaire, the student roster and identification of students with disability	≥ 2 teachers per school	≥ 4 for baseline study	30 minutes
Classroom observation	≥ 2 classes per school	Grades 1–5 (≥ 4 classes) for baseline study and Grades 3-6 for endline study	45 minutes per class
School supervisor survey	1 supervisor per school		30 minutes
Students survey grades 1–6[4]	Minimum of 20 students per school		15 minutes
Student test – grade 1	5 students per grade (min. 20 students)	Mathematics, Bahasa Indonesia and Raven test	30 minutes per subject per student
Student test – grade 2		Mathematics, Bahasa Indonesia and Raven test	
Student test – grade 3		Mathematics, Bahasa Indonesia and Raven test	
Student test – grade 4		Mathematics, Bahasa Indonesia and Raven test	45 minutes per subject per students
Student test – grade 5		Mathematics, Bahasa Indonesia and Raven test	
Student test – grade 6		Mathematics, Bahasa Indonesia and Raven test	

Parent survey	Minimum of 20 parents		45– 60 minutes
School observation			15 minutes

Notes: The Raven test is an established non-verbal test of fluid intelligence.

Some adjustments were made to the baseline for the SIPPI–2018 West Nusa Tenggara endline study. During the endline data collection process, most of the instruments were in a digital-based format with only a few in a paper-based format. The digital-based program was built on a tablet under an Android platform using the SurveyCTO data collection program. Some adjustments to the questions were made to be able to capture the effect of the intervention and difference in the timing of the survey. For teachers, this endline study tracked the panel respondents from the baseline study in 2016. Teachers participating in the endline should have also been involved in the Guru BAIK and Gema Literasi interventions.

Pre-printed lists of the schools and the panel respondents (students, teachers, school principals, school supervisors and parents) were prepared to capture the profile of the respondents in the 2016 baseline survey. The pre-printed lists were used to confirm the availability of the panel respondents by making initial phone-calls with the principals. Thus, the pre-printed lists were to help the teams identify all the endline survey respondents.

2.2 SURVEY IMPLEMENTATION

Under the supervision of INOVASI's monitoring, evaluation, research and learning (MERL) team, the Regional Economic Development Institute (REDI), an independent research institute based in Surabaya, carried out the survey. Implementing the survey involved two phases: training the enumerators and collecting the data.

Training the enumerators

Prior to data collection, a training program was held in Mataram, West Nusa Tenggara, at the Grand Legi hotel from 30 July to 5 August 2018. Overall, 86 prospective enumerators took part and 80 were assigned to field survey teams with six being retained as reserves. Among these enumerators, 53 (62 per cent) were from West Nusa Tenggara and 33 (39 per cent) were from elsewhere. Local enumerators are essential because they understand the local language, have local knowledge and know how to access further information about the area.

The training program included an introduction to INOVASI, our child protection policy and the endline studies for Guru BAIK and Gema Literasi. Participants were also trained on the survey instruments, targeting respondents, sampling techniques, and data collection arrangements and management. The course was conducted over six days and applied several training methods, such as lectures, practical work, demonstrations and discussions, depending on the content. Most of the survey instruments were delivered through lectures. All prospective enumerators needed to have the same understanding of each question in the survey instrument. Each question has a specific required value, period of time or focus that all enumerators need to respect.

Another method used in the training was having live interviews with 'dummy' respondents. The REDI team invited school supervisors, principals, teachers, students and parents from the nearest schools to the training venue. These dummy respondents were not part of the targeted respondents for the main data collection but these live sessions gave the enumerators practical experience in delivering the instruments to the targeted respondents.

Collecting the data

REDI deployed 20 field teams: ten teams were deployed to North Lombok and ten teams to Sumbawa. Each field team consisted of a supervisor and three enumerators. The teams' target was to complete collecting data at a school within three days.

Unfortunately, earthquakes hit West Nusa Tenggara just as the field teams arrived at the research sites. North Lombok was hit by a 7.0 Richter scale (SR) earthquake on 5 August 2018 at 7.46 in the evening (local time). The earthquake caused injuries and the death of hundreds of people, massive destruction of buildings and infrastructure (roads, bridges, official buildings, houses), including the buildings of most partner schools in North Lombok. The conditions in North Lombok were unsafe and most people were moved to refugee camps. Most damage occurred in Kayangan, Tanjung, Bayan and Pemenang sub-districts which are part of the location for this endline survey. Two research base camps were reported to be damaged and the survey material was buried in the ruins. The survey was then put on hold.

At a later stage, Sumbawa was also hit by a 7.0 SR earthquake causing damage to infrastructure and buildings in the region. The team took a similar action to North Lombok, taking a temporary break to assess whether the data collection process could feasibly be continued. At that time, the Governor of West Nusa Tenggara made an official announcement about terminating the school process in the province for an undetermined period. INOVASI Jakarta also suggested the team in Sumbawa should withdraw since several small-scale earthquakes had occurred. The data collection process in Sumbawa was two thirds of the way through at this stage.

The overall data collected for the endline survey in Sumbawa district is summarised in table 2.2.

Table 2. 2: Total number of actual respondents for each instrument

No	Respondent type	Number of respondents
1a	Principal	36
1b	School observation	36
2a	Teacher[5]	94
2b	Classroom observation	89
3	School supervisor	7
4	Student survey	474
5	Parent survey	474
6	School committee	N/A

The recapitulation shows only data collected from Sumbawa as no data was collected in North Lombok (50 school sites). Based on the percentage of complete data, the data collection status can be classified into three categories. The process is considered 'completed' when all the required data is collected and the team had achieved this in 32 schools (64 per cent). Meanwhile, in 12 schools (24 per cent), the data was incomplete at the time of the time the team was withdrawn. The third category of 'not visited' means that the team did not begin collecting data at all and six schools (12 per cent) fell into this category.

We should note that while the endline survey was conducted at the beginning of the academic year, the baseline survey was conducted at the end of the semester. This timing difference might affect some indicators examined in this study and therefore comparisons between the baseline and endline results and their interpretations can only be made with caveats.

[5] Among these teachers, approximately 87 per cent taught literacy and 88 per cent taught numeracy during the observations.

3. MAIN FINDINGS

This chapter presents the main findings from the endline study and some comparisons with the baseline results. There are two main sections. The first examines the effects indicators reflecting improvements in students and teachers' performance. To confirm these findings, we also explore the perspectives of other stakeholders, including school principals, supervisors and parents. However, we should note that Guru BAIK did not directly involve parents. The second section identifies factors that might correlate with students' performance in literacy and numeracy based on the baseline and endline data.

3.1 EFFECT

The main indicators relating to students in the Guru BAIK intervention are: students' performance in literacy and numeracy; and their attitudes towards literacy and numeracy learning. We begin by looking at students' performance and examining whether the Guru BAIK intervention led to better scores. A procedure was developed to ensure that the results truly reflect the effect of Guru BAIK, especially for the results relating directly to students' performance. We then review the descriptive statistics on students' enthusiasm for learning and their attitudes towards Bahasa Indonesia and mathematics, the two key subjects in the pilot program.

Students' performance in literacy and numeracy

The baseline and endline surveys assessed students' performance in literacy and numeracy. Students' responses to the test items were analysed using a two-parameter logistic item response theory (IRT) model that gives information on students' ability based on the probability of them responding correctly to items at different difficulty levels and on the items' discrimination power.

The test results were then equated to make them comparable, conducted a psychometrician.[6] Equating is the process of linking two or more tests that have a similar measurement target but different structures, compositions and psychometric parameters. The averages and standard deviations of the two tests' scores may vary due to these differences. For example, a test with more items tends to lead to higher scores compared to one with fewer items. Anchor item parameters' estimations were used as part of the equating process. The procedure is critical to ensure that the scores from different tests have the same meaning over time.[7]

Following the equating administration, the next procedure is statistically adjusting the net increase or decrease of the treated group's average scores. Similar procedure for control group with the average scores of the group in the baseline and endline surveys were held the same (see figures 3.3 and 3.4). With this procedure, the net effect can be clearly counted, by identifying gaps between the average scores of the treated and the control groups.

Equating was necessary in our case because the time period between the baseline and endline surveys was two years. Students who were in grade one during the baseline survey were in grade three when the endline survey was conducted and consequently they received different literacy tests. A direct comparison between the results of the grade three and grade one tests would not illustrate the actual change in scores.

Figures 3.1 and 3.2 illustrate the levels of difficulty of the Bahasa Indonesia and mathematics tests for different grades. Item difficulty is an important concept in psychometry and reflects how likely a person provides an incorrect response towards a question. It is calculated by dividing the number of people who got wrong answers by the total number of people responding to the question, followed by a normalisation (mean equals to 0 and standard deviation equals to 1). Some questions needed to be equated using the anchor items to ensure

[6] The psychometrician is one of MERL consultant panel members. He is a lecturer in Faculty of Psychology of Gadjah Mada University with extensive experience in conducting psychometric analysis.
[7] Dorans, Neil J., Moses, Tim P. and Eignor, Daniel R. 2010. *Principles and Practices of Test Score Equating*. Accessed through internet on April, 2019, https://www.ets.org/Media/Research/pdf/RR-10-29.pdf.

comparability. As the figures show, the tests for those in the higher grades were more difficult meaning that students in these grades were assessed according to the expectations of their abilities.

Figure 3. 1: Difficulty level: Bahasa Indonesia

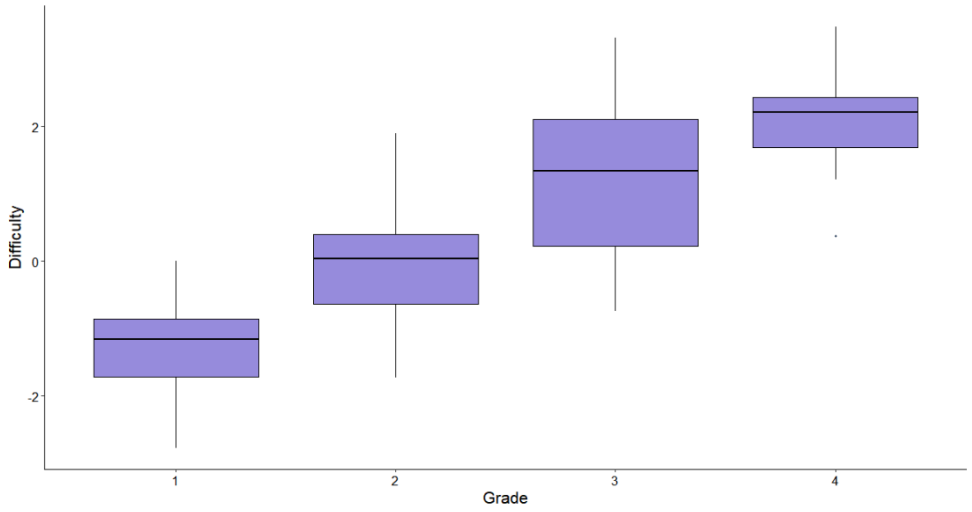
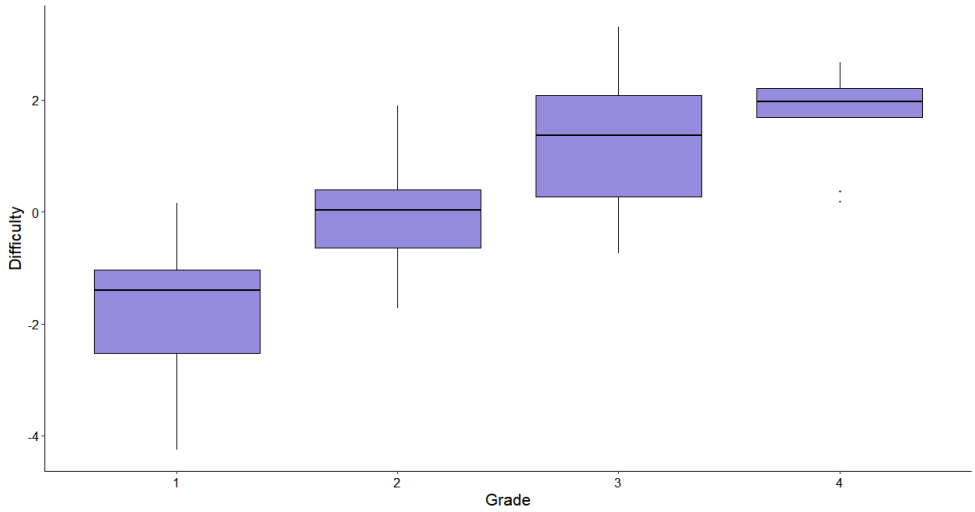


Figure 3. 2 Difficulty level: mathematics



A generalised analysis of covariance (g-Ancova) takes the endline score as the dependent variable and the treatment group variable as the independent variable. Other variables such as baseline scores, gender and class are included as control variables. The use of g-Ancova provides more accurate results in experimental research using non-randomised and non-equivalent designs because the technique accommodates several limitations, including potential interactions and data heterogeneity (Widhiarso 2018). The following sub-sections present the findings on students performance in literacy and numeracy.

Literacy

The literacy assessment tools for grades one and two consisted of 24 and 25 items respectively, and twelve of these are the anchor items. The test included: letter, word and sentence recognition; vocabulary and word

usage; and explicit information retrieval from short passages. In terms of the cognitive domain, these tests only assess students' lower order thinking skills (knowing).

Each test for grades three and four consisted of 24 to 28 questions and about half of these were the anchor items. The tests cover writing and reading skills: vocabulary and word usage; grammar and punctuation; text organisation; focusing on and retrieving explicitly-stated information; making straightforward inferences; interpreting and integrating ideas and information; evaluating and critiquing content and textual elements. The tests for the higher grades also assessed higher order thinking skills (applying, reasoning and creating). Figure 3.3 shows that the net increase in students' literacy score was 1.2 points.[8] This is one indication that the intervention was beneficial for the students.

Figure 3. 3: Students' average literacy score (out of 100)

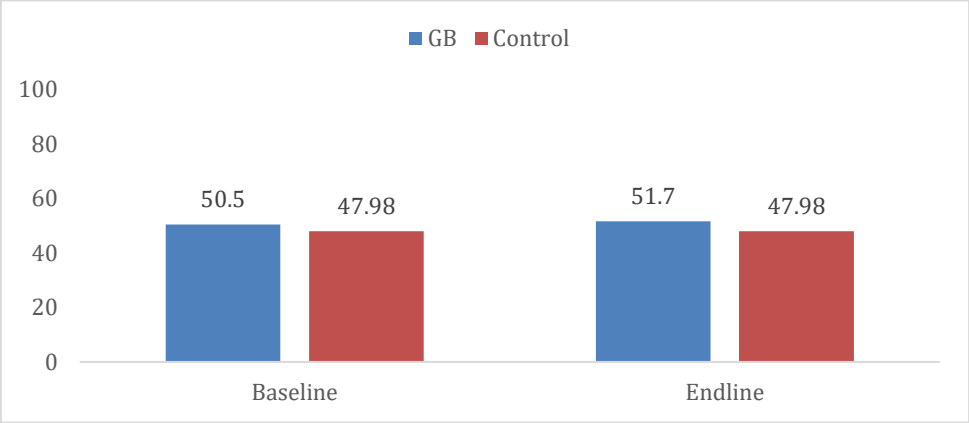


Table 3. 1: Comparative effects on students' average literacy score of the Guru BAIK and the Gema Literasi interventions, by gender

Category	Guru BAIK					Gema Literasi				
	Coefficient	SE	SD	p.value	Effect Size	Coefficient	SE	SD	p.value	Effect Size
All	1.95	0.77	2.53	0.0056	0.2076	0.578	0.789	11.38	0.2318	0.0615
Boys	2.32	1.09	2.12	0.017	0.2466	0.789	1.09	15.72	0.2355	0.0839
Girls	1.61	1.09	1.49	0.0689	0.1715	0.382	1.13	16.3	0.36785	0.0407

Notes: SE = standard error; SD = standard deviation; p.value = calculated probability

Table 3. 2: Comparative effects of on students' average literacy score of the Guru BAIK and the Gema Literasi interventions, by grades

Grade	Guru BAIK					Gema Literasi				
	Coefficient	SE	SD	p.value	Effect Size	Coefficient	SE	SD	p.value	Effect Size
1	2.28	0.17	1.28	0	1.78	0.03	0.3	2.54	0.46	0.01
2	1.94	0.11	0.93	0	2.09	0.01	0.39	3.01	0.49	0.01
3	1.07	0.15	1.32	0	0.81	0.52	0.31	2.51	0.05	0.21
4	-0.63	0.25	1.9	0.01	-0.33	0.1	0.54	3.2	0.43	0.03

Notes: SE = standard error; SD = standard deviation; p.value = calculated probability

[8] The average literacy scores of the control group during the baseline and endline studies were similar due to the adjustment using the equating procedure.

As illustrated in table 3.1, using the equated IRT score and g-Ancova procedure, the average increase in the literacy score due to the Guru BAIK intervention was 1.95 which is statistically significant. When the sample was disaggregated into boys and girls, the effect of the intervention was higher for boys than for girls (scores increased by 2.32 and 1.61, respectively). Compared to the Gema Literasi intervention, Guru BAIK led to a greater improvement in students' performance in literacy.

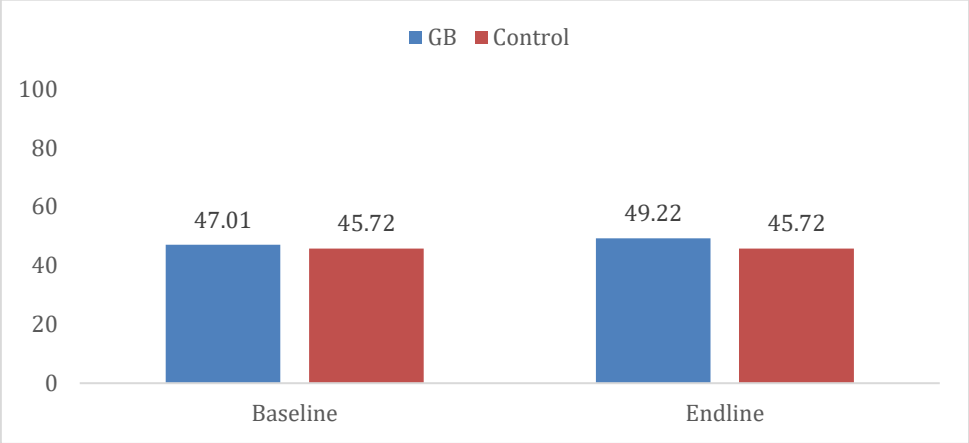
Table 3.2 provides the breakdown based on grades.[9] As shown, the intervention had the most positive effect among grade one students (2.28). The effect was progressively decreasing at higher grades, and the effect of the intervention was negative among grade four students. At early grades, Guru BAIK was more beneficial for students' performance in literacy compared to the Gema Literasi intervention, except in the case of grade four students.

Numeracy

The numeracy assessment tools for grade one consisted of 25 items to assess how students recognise, classify, order or compute whole numbers. In terms of the cognitive domain, all items for grade one were classified as lower order thinking skills (knowing). Starting from grade two, items assessing higher order thinking skills were included. Two out of the 30 items for grade two students assessed the second hierarchy cognitive domain of applying knowledge. Similarly, three out of the 27 items for grade three students also assessed the skill of applying knowledge. Items to assess reasoning ability were included in grades four and five. About half of all items in each grade served as the anchor items that enable us to make comparisons across grades. Similar to the literacy scores, students' numeracy scores reflect each student's latent ability obtained from a two-parameter logistic item response theory model.

The net increase in students' average numeracy score was 2.21 – higher than the net increase in literacy score, as shown in figure 3.4.[10] This indicates that the Guru BAIK intervention led to improvements in students' average numeracy score.

Figure 3. 4: Students' average numeracy scores (out of 100)



In general, the result from the IRT score and g-Ancova procedure shows that Guru BAIK led to an average net increase of 2.92 in students' numeracy scores and this effect is statistically significant. In contrast to our findings for the literacy scores, the Guru BAIK intervention only had a significant effect on numeracy scores for boys, with an estimated coefficient of 3.89 (table 3.3). Disaggregating the sample based on grades, the intervention had the highest effect among grade one students and this is consistent with the findings on students' literacy. Meanwhile, in general the intervention had the lowest effect on grade three students (table 3.4).

[9] This breakdown is provided according to students' grades during the baseline study. Those who were in grades five and six at that time did not participate in the endline survey since they were no longer at the primary schools.
[10] The average numeracy scores for the control group during the baseline and endline studies were similar due to the equating procedure.

Table 3. 3: Effect of the Guru BAIK intervention on students' average numeracy scores, by gender

Category	Coefficient	SE	SD	p.value	Effect size
All	2.92	1.21	2.42	0.0077	0.251
Male	3.89	1.65	2.35	0.0093	0.334
Female	1.97	1.76	1.12	0.1314	0.169

Notes: SE = standard error; SD = standard deviation; p.value = calculated probability

Table 3. 4: Effect of the Guru BAIK intervention on students' average numeracy scores, by grade

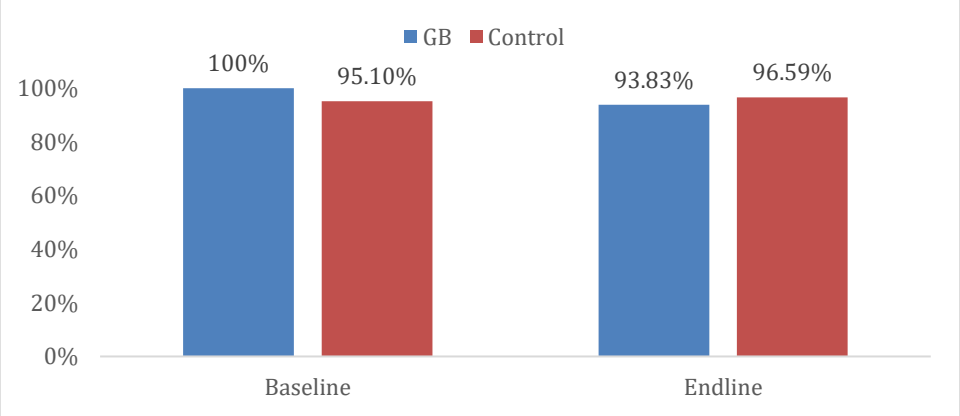
Grade	Coefficient	SE	SD	p.value	Effect size
1	5.06	0.94	4.52	0.00	1.12
2	2.93	0.59	2.48	0.00	1.18
3	0.67	0.42	2.40	0.05	0.28
4	1.26	0.86	4.64	0.07	0.27

Notes: SE = standard error; SD = standard deviation; p.value = calculated probability

Students' attitudes

Students were asked about their enthusiasm for going to school[11] during the baseline and endline surveys. Figure 3.5 illustrates that while initially all students in Guru BAIK group were enthusiastic about going to school, in the endline, the percentage of students giving similar responses had gone down (by almost 6 per cent). On the other hand, there was an increase in the percentage of students who were excited about going to school among the control group. The differences in the students' responses between Guru BAIK and the control groups in the endline survey were statistically significant. One possible explanation for the lack of enthusiasm for going to school among the Guru BAIK group could be the teachers' greater use of active learning approaches that put students at the centre of the learning activities. This shift in roles might be perceived as more demanding by the students. However, the relationship between students' perceived difficulty with a subject and their motivation is complex and depends on many factors (Cuff 2017). Further investigation is required to confirm whether more demanding lessons would lead to less enthusiasm for school.

Figure 3. 5: Students who were excited about going to school (percentages)

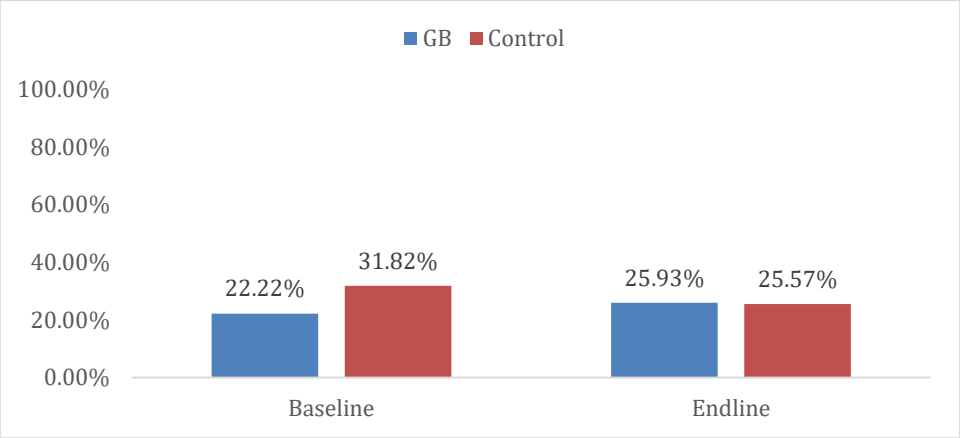


Students' attitudes towards a subject might influence their performance in the subject. To capture this, students were asked about their favourite subject at school. Figure 3.6 shows that the proportion of students that chose Bahasa Indonesia as their favourite subject among Guru BAIK group increased from the baseline to the endline

[11] In practice, students' enthusiasm for going to school reflects not only their attitude to learning but also many different elements such as their eagerness to interact with school friends, whether they like their teachers or principals, their resilience in waking up early, and so on.

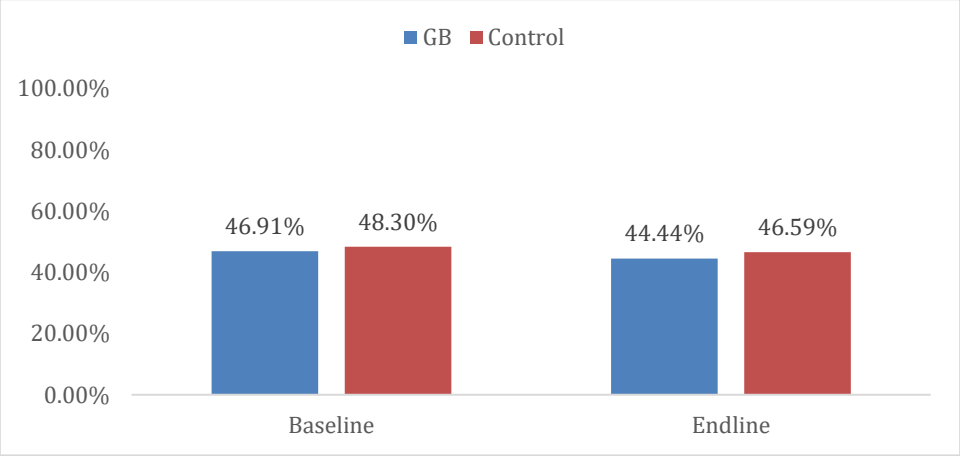
studies (from 22.22 per cent to 25.93 per cent). In contrast, the proportion of students that chose the subject as their favourite in the control group decreased from 31.82 per cent to 25.57 per cent. These two opposite effects led to a net change difference of 9.96 per cent for Guru BAIK group.

Figure 3. 6: Students choosing Bahasa Indonesia as their favourite subject (percentages)



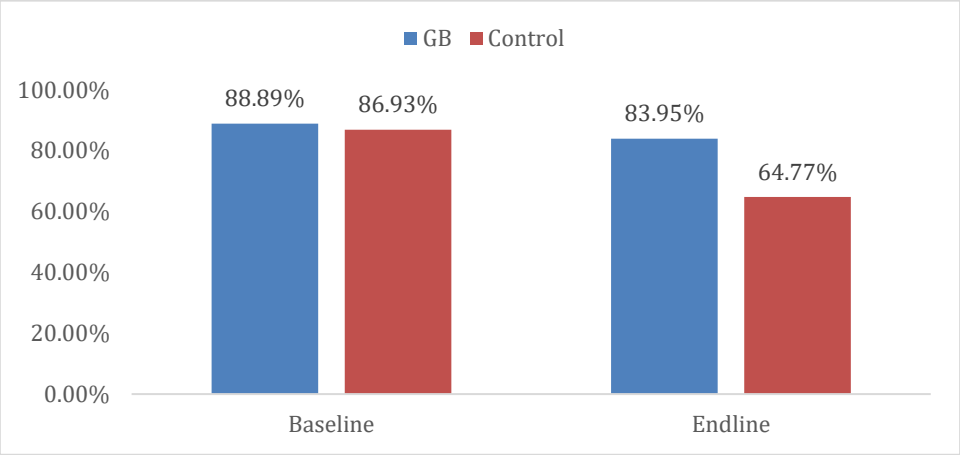
In both the Guru BAIK and the control groups, students' attitude towards mathematics tended to be less enthusiastic in the endline survey, as shown in figure 3.7. This is indicated by fewer students choosing mathematics as their favourite subject (down by 2.47 per cent and 1.71 per cent for the Guru BAIK and control groups, respectively). This might be due to the increasing level of difficulty in mathematics based on students' perceptions. Further qualitative inquiries might help to confirm this, as the relationship between perceived difficulties and attitude towards a subject is inconclusive (Cuff 2017).

Figure 3. 7: Students choosing mathematics as their favourite subject (percentages)



Students' attitude towards learning can also be confirmed from the parents' perspectives. Parents were asked their views on their children's efforts to achieve good grades. Similarly, the findings reflected that fewer parents, among both the Guru BAIK and the control groups, believe their children make an effort to get good grades (figure 3.8). The reduction among the control group was more significant than in Guru BAIK group (around 22 per cent versus 5 per cent, respectively), leading to a net increase for the Guru BAIK group of approximately 17 per cent. In the endline, the prevalence of parents who had positive views on their children's efforts was higher for the Guru BAIK group and the differences were statistically significant.

Figure 3. 8: Parents who believe their children made a great effort to better learning results (percentages)



Information about parents' perceptions of the benefits of school for childrens' development was also collected. The report focuses on five aspects related to this: literacy development (figure 3.9); numeracy development (figure 3.10); making friends (figure 3.11); identifying role models (figure 3.12); and improving behaviour (figure 3.13). More parents viewed schools as useful for the first four aspects among both the Guru BAIK and control groups. However the net effects for literacy development and numeracy development for the Guru BAIK groups were negative (approximately -16 per cent and -3.55 per cent, respectively) due to a larger increase in the control groups. The differences in parents' responses between the Guru BAIK and control groups with regard to these two aspects were statistically significant in both the baseline and endline surveys. Moreover, despite more parents having positive views on the usefulness of schools for making friends and identifying role models, the levels were still remarkably low. Statistically significant differences were only found during the baseline survey in parents' responses on the role of school in helping students identify role models.

The negative net change difference for the Guru BAIK group when it came to parents' perceptions of the importance of school in improving literacy and numeracy deserves more attention than parents' perceptions of other aspects. Nevertheless, it should be noted that the Guru BAIK approach did not involve parents in its interventions. Thus, there is no control or engagement with parents, resulting in lower level of information and communication between parents and teachers. With the current available data, it is not possible to identify why the control group had stronger parents' perspective on this issue. Possibly, the control group had activities that involved parents, which did not take place in Guru BAIK schools. In addition, school principals were also not involved heavily in the Guru BAIK pilot. Another possible assumption is teacher rotation, which was confirmed during the endline study. Parents may provide their perceptions based on the current teachers which are not necessarily always the same as the Guru BAIK teachers. Another assumption worth investigating further is that parents may perceive they have a role to play in improving literacy and numeracy outside of the school. These assumptions require further data support.

Figure 3. 9: Parents who viewed schools as useful for students' literacy development (percentages)

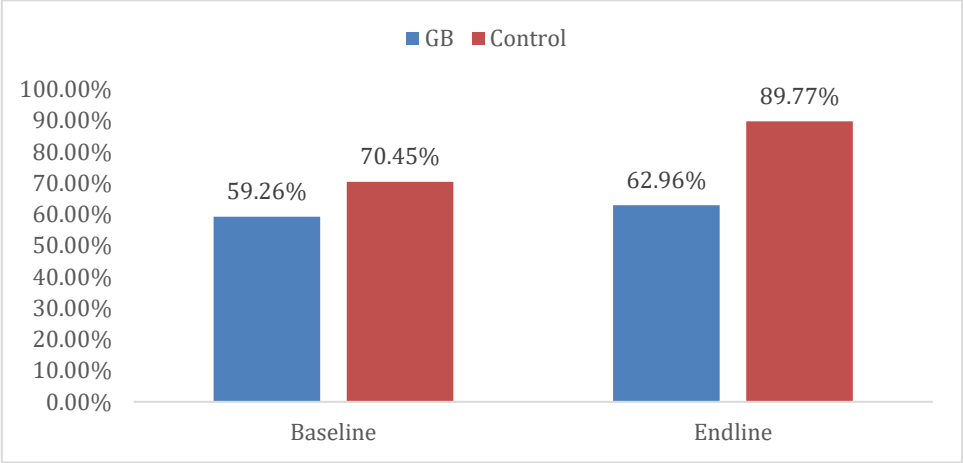


Figure 3. 10: Parents who view schools as useful for students' numeracy development (percentages)

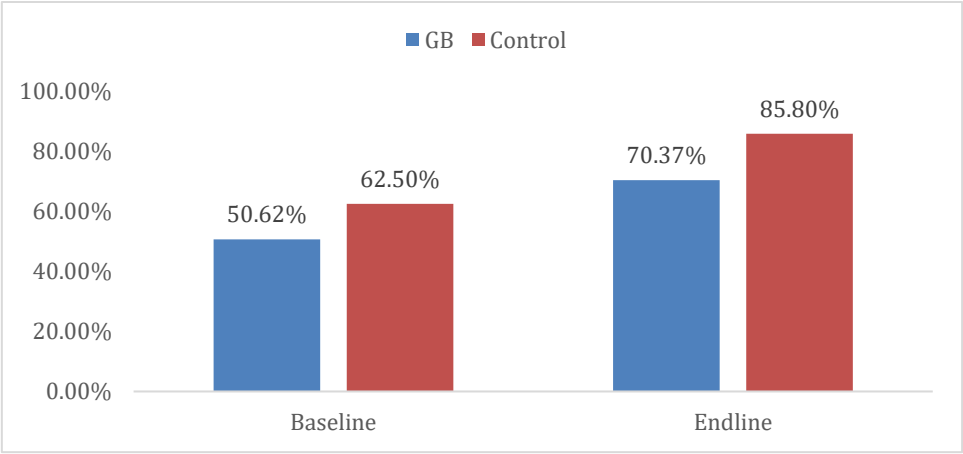


Figure 3. 11: Parents who viewed schools as useful for making friends (percentages)

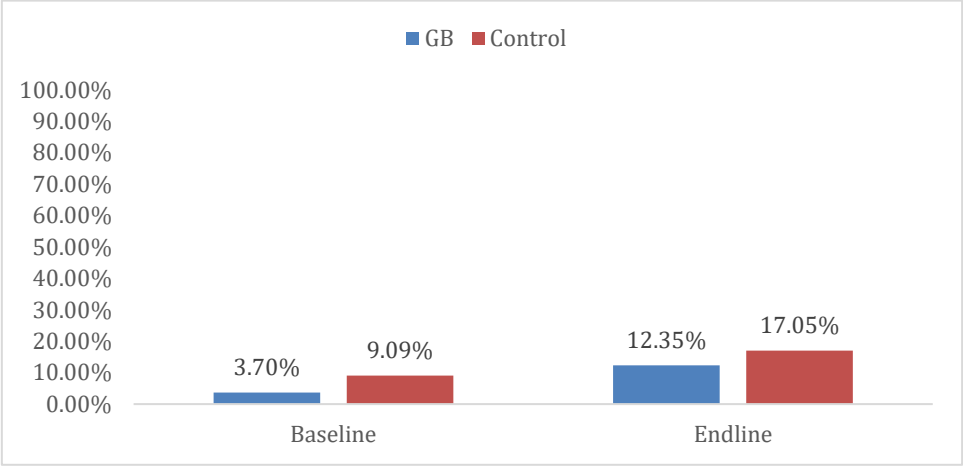
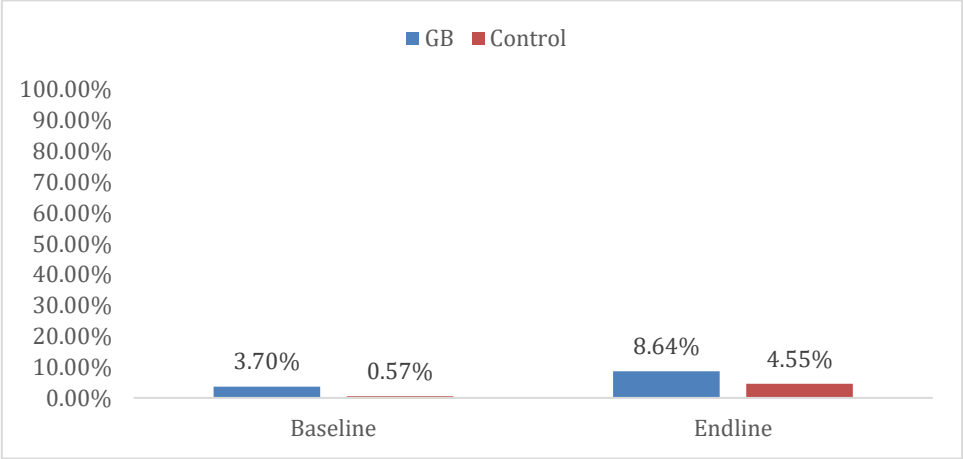
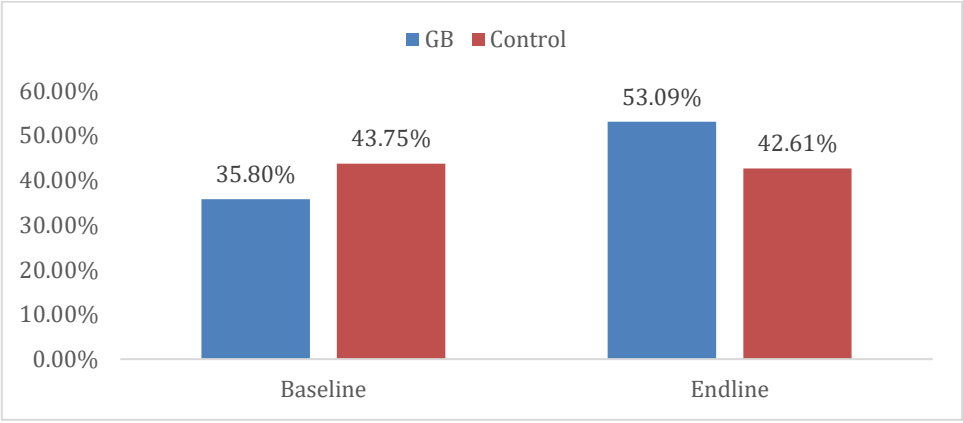


Figure 3. 12: Parents who viewed schools as useful for identifying role models (percentages)



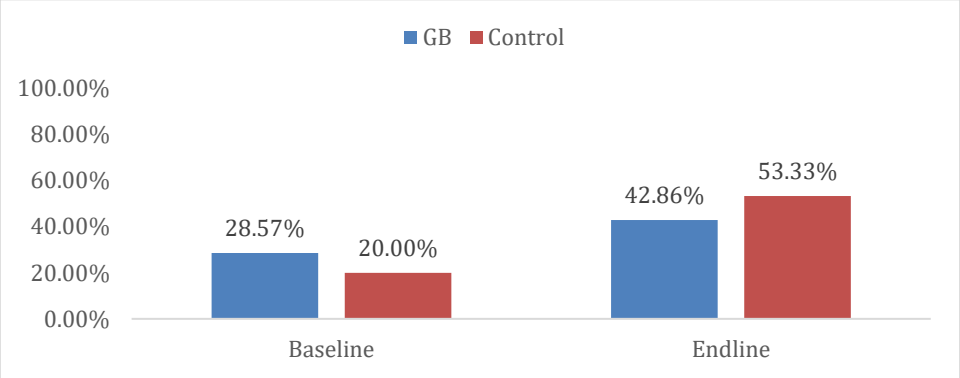
Meanwhile, opposite trends were identified in the case of parents who viewed schools as useful to improve behaviour. A surprising increase of almost 20 per cent was recorded among the Guru BAIK group while among the control group, there was a 0.69 per cent decrease.

Figure 3. 13: Parents who viewed schools as useful to improve behaviour (percentages)



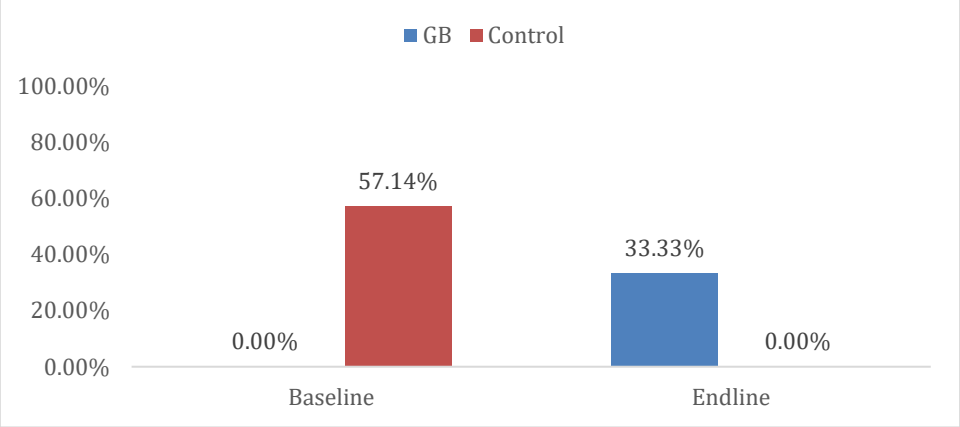
Aside from parents, the perspectives of school principals and supervisors can also indicate improvements in students' learning outcomes (figure 3.14). In both the Guru BAIK and the control groups, more school principals were satisfied with students' learning outcomes in the endline survey, with a more significant change in the case of the control group (around 33 per cent).

Figure 3. 14: School principals satisfied with students' learning outcomes (percentages)



The trends were different for Guru BAIK and the control groups with regard to supervisors' satisfaction with students' learning outcomes. In the Guru BAIK group, none of the school supervisors was satisfied with students' learning outcomes in the baseline survey but around 33 per cent were satisfied in the endline survey. In contrast, more than half of the supervisors in the control group were satisfied with students's learning outcome in the baseline but this decreased to none in the endline (figure 3.15).

Figure 3. 15: School supervisors satisfied with students' learning outcomes (percentages)



Teaching practices

The first teaching practice we examine is teachers' use of active instruction in the learning process. Activities that can be classified as active instruction include: reading aloud; demonstrating or lecturing; and discussing or practising drills (Bruns and Luque, 2014 p.106). Teachers are using active instruction if they perform at least one of these activities. Figure 3.16 shows that teachers in both the Guru BAIK and the control groups increasingly used active instruction. The percentage of teachers using these methods of teaching in the two groups increased by around 11 per cent and 10 per cent, respectively, leading to the net effect of 1 per cent increase for the Guru BAIK group. At the same time, fewer teachers used passive instruction methods, such as monitoring copying and doing in-class assignments (figure 3.17).

Figure 3. 16: Teachers using active instruction (percentages)

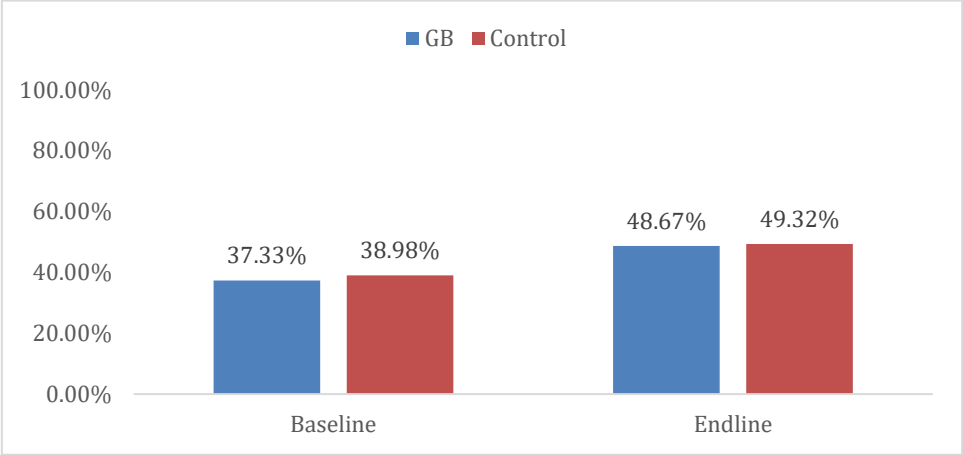
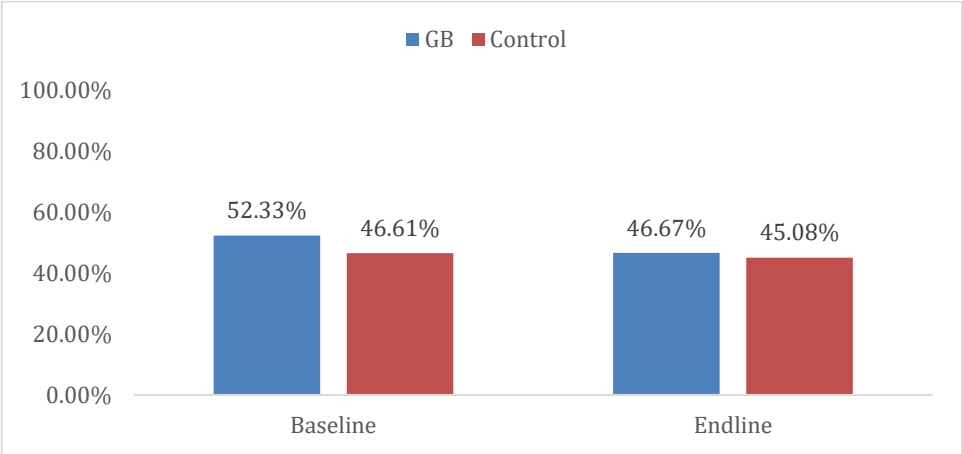
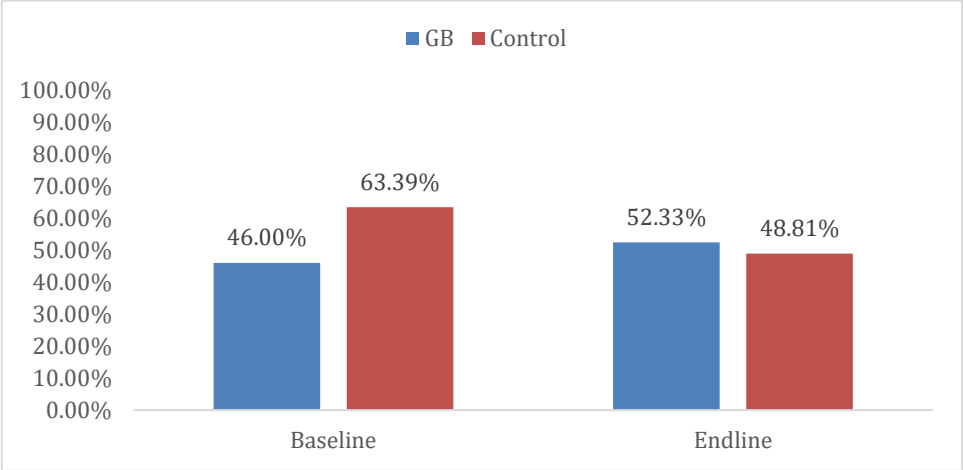


Figure 3. 17: Teachers using passive instruction (percentages)



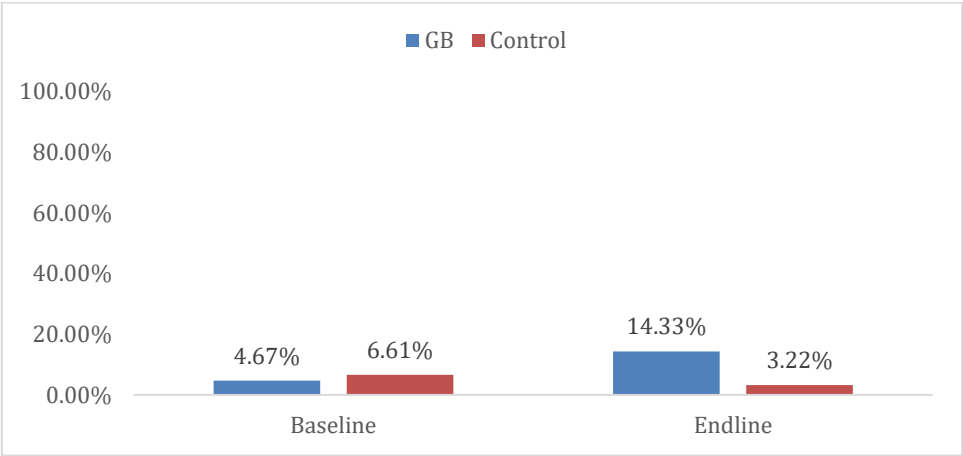
One way of evaluating the effectiveness of a learning process is to establish whether students were engaged or not. Enumerators recorded the number of students who did not seem to pay attention or engage with the learning activities in the classroom. The learning process is considered engaging if all students pay attention during the process. As shown in figure 3.18, while in Guru BAIK group more classes with engaged students were observed (increase of 6.33 per cent), the trend in the control group was the opposite (decrease of around 15 per cent). Therefore, the net increase for Guru BAIK group was nearly 21 per cent.

Figure 3. 18: Classes with engaged students (percentages)



Classroom observations also captured teachers' use of learning media. Figure 3.19 indicates that the use of learning media in both the Guru BAIK and the control schools was still limited to below 15 per cent. While there was a positive trend in the use of learning media among Guru BAIK schools (significant increase of almost 10 per cent), the control schools showed a negative trend in the endline survey (decrease of 3.39 per cent). This means the net effect for the use of learning media among the Guru BAIK group was around 13 per cent.

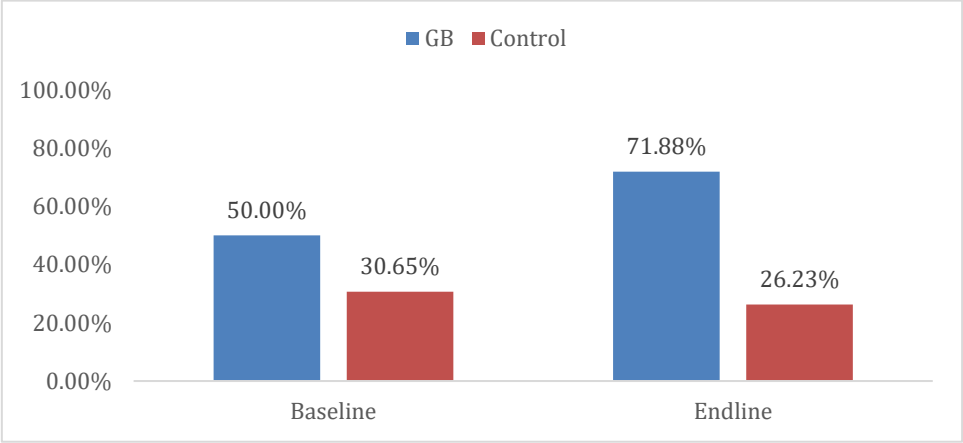
Figure 3. 19: Teachers using learning media (percentages)



In the surveys, teachers were asked about the types of assignments they gave their students. Three assignments that are considered effective in improving students' literacy levels are: reading from other sources; putting things they learned into practice; and writing about their opinions on subjects. The survey examined the percentage of teachers using these activities on a daily basis.

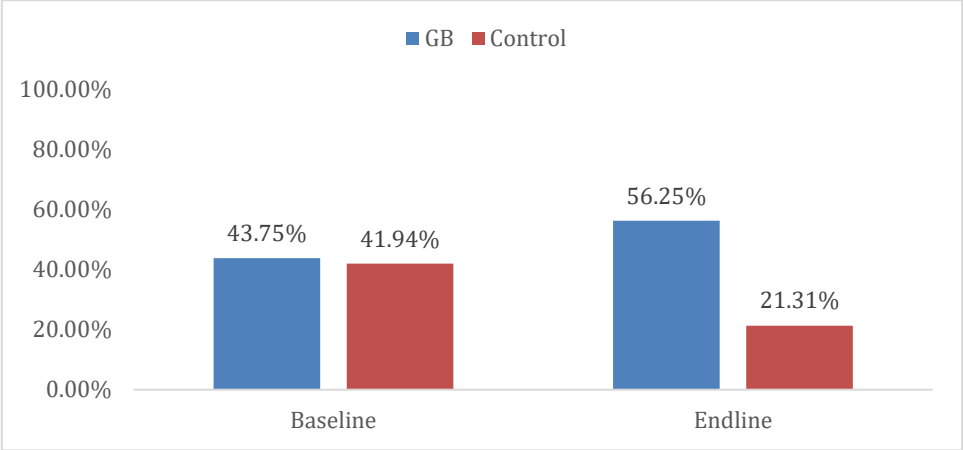
Figure 3.20 shows that the percentage of teachers asking students to read from other sources every day increased by around 22 per cent among the Guru BAIK group. In contrast, the percentage decreased by 4.42 per cent from the baseline to the endline surveys among the control group. This led to a net increase of more than 26 per cent for the Guru BAIK group.

Figure 3. 20: Teachers asking students to read from other sources every day (percentages)



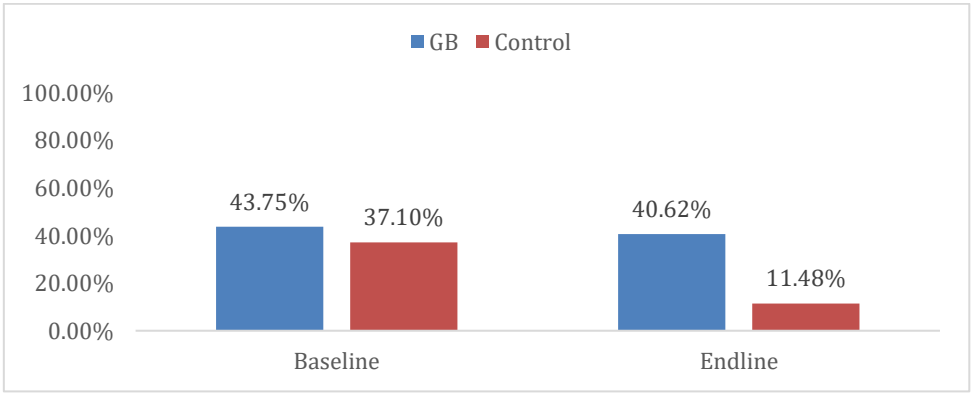
A similar trend was found with regard to teachers asking students to put what they have learned into practice. While the percentage of teachers setting this assignment increased by almost 13 per cent among the Guru BAIK group, the percentage among the control group decreased significantly by around 21 per cent. The net change difference as a consequence of these two effects was 33 per cent for the Guru BAIK group.

Figure 3. 21: Teachers asking students to put what they have learned into practice (percentages)



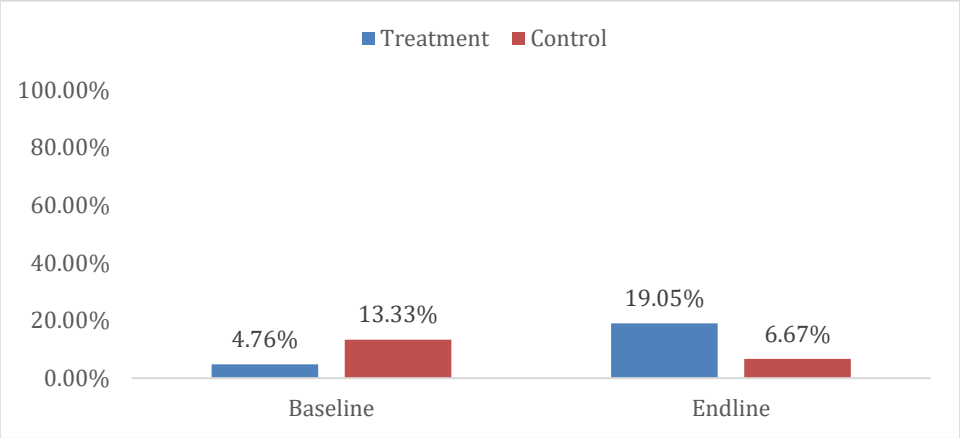
In contrast, the percentage of teachers asking students to write their opinions about subjects every day decreased for both groups as shown in figure 3.22. However, as the percentage decrease in the control group was more significant (27 per cent compared to 3.13 per cent), this resulted in a positive net effect for the Guru BAIK group of more than 22 per cent.

Figure 3. 22: Teachers asking students to write their opinions about subjects every day (percentages)



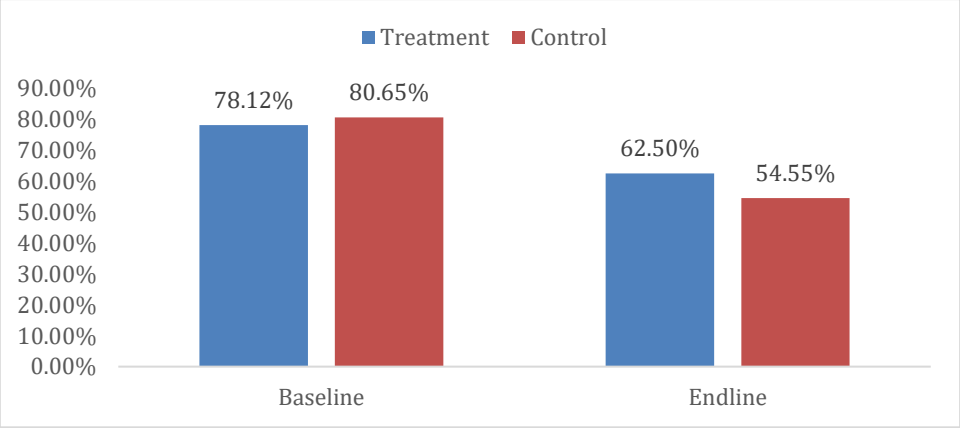
We explored school principals' perceptions of teachers' workloads and found that generally only a few (less than 20 per cent) believed that teachers had too much work. However, among teachers in the Guru BAIK group, there was an increase of more than 14 per cent between the baseline and the endline while among the control group there was a decrease of almost 7 per cent, resulting in a net increase for the Guru BAIK group of almost 21 per cent (figure 3.23).

Figure 3. 23: School principals who believe that teachers had too big a workload (percentages)



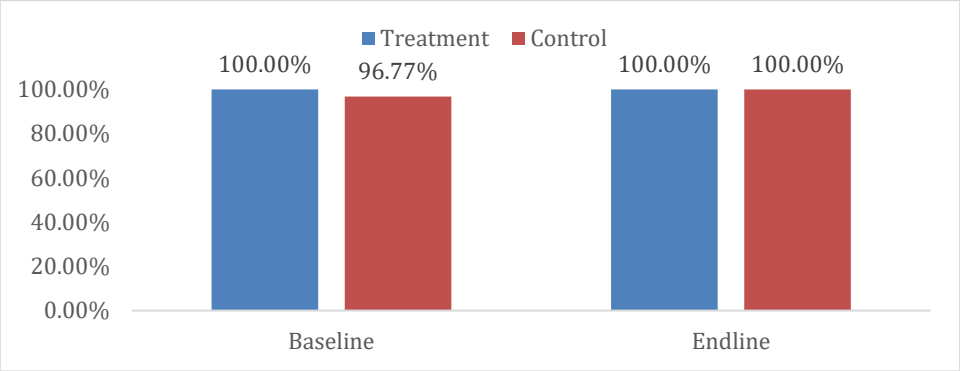
The study also examined teaching practices through indicators reflecting the use of workplans. We found that fewer teachers developed and were able to show workplans in the endline survey for both for the Guru BAIK and control groups. Nevertheless, the decrease was larger among the control group, leading to a net positive change difference for the Guru BAIK group of more than 10 per cent. In addition, the Guru BAIK group still performed better in the endline survey (figure 3.24).

Figure 3. 24: Teachers who developed workplans and were able to show them (percentages)



A checks and balances process such as supervision by school principals, supervisors, committees and other stakeholders can help ensure that teaching practices achieve their intended outcomes. In the surveys teachers were asked about the usefulness of this supervision. In both the baseline and endline surveys, all the teachers in Guru BAIK group considered the feedback from such supervision as useful (figure 3.25). The control group was similarly positive about this feedback.

Figure 3. 25: Teachers who considered feedback from supervision as useful (percentages)



The Guru BAIK intervention should lead to more awareness among teachers about their need for professional development. We compared teachers' training and professional development activities in the Guru BAIK group and the control group, including their participation in teachers' working group (KKG) meetings. In both the baseline and endline surveys, more teachers from the Guru BAIK group participated in teachers' working group meetings (figure 3.26), with statistically significant differences in both the baseline and endline surveys. Another notable finding was the increasing trend of teachers participating in these meetings (increases of 3.12 per cent and 1.61 per cent for Guru BAIK and the controlled groups, respectively). Figure 3.27 provides information on the topics discussed during teachers' working group meetings and shows a higher frequency of discussions about students' learning outcomes based on the information in the endline survey for both Guru BAIK and the control groups. Nevertheless, the increase in the control group was higher, leading to a negative net change difference of -11 per cent for the Guru BAIK group. The topic with the highest net change difference was preparing students' evaluations.

Figure 3. 26: Teachers participating in teachers' working group meetings (percentages)

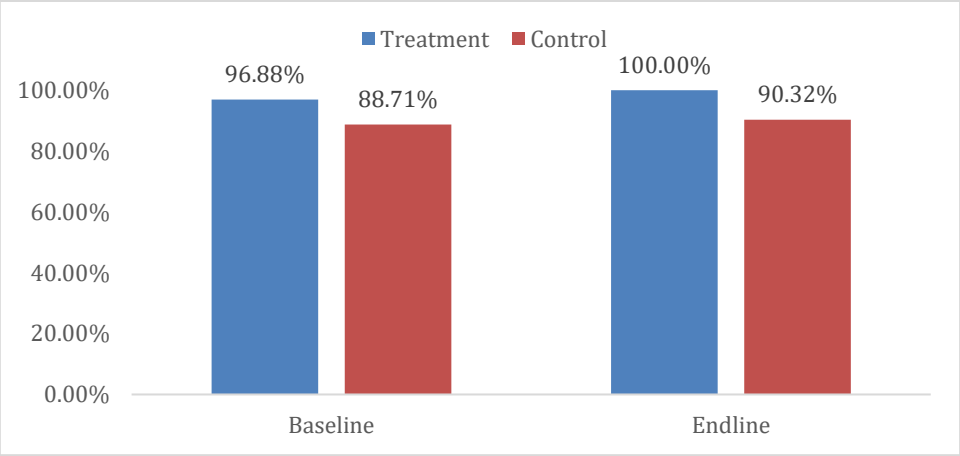
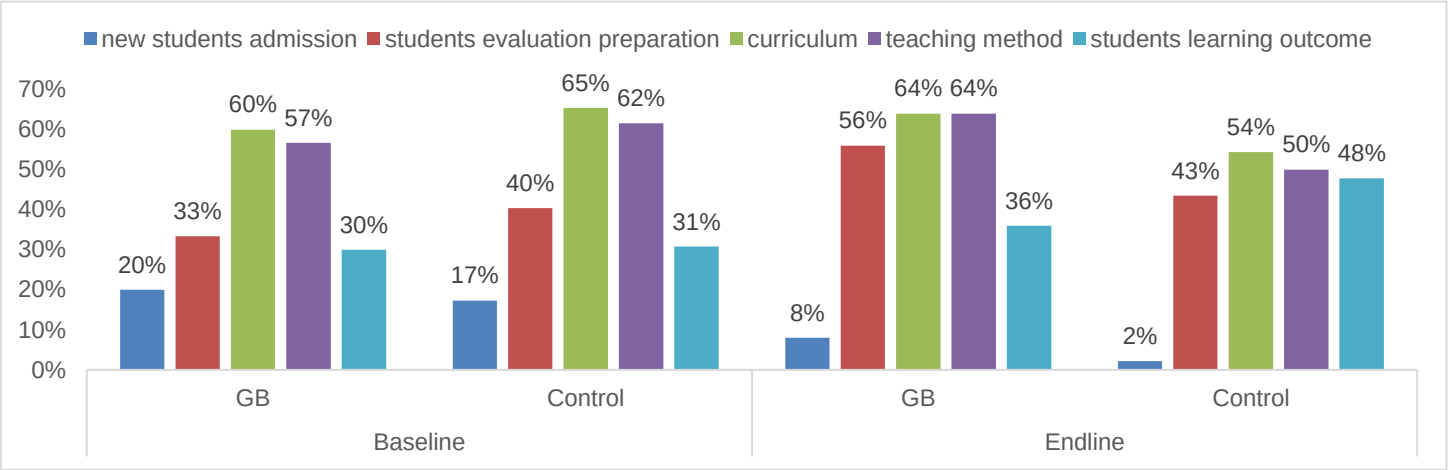
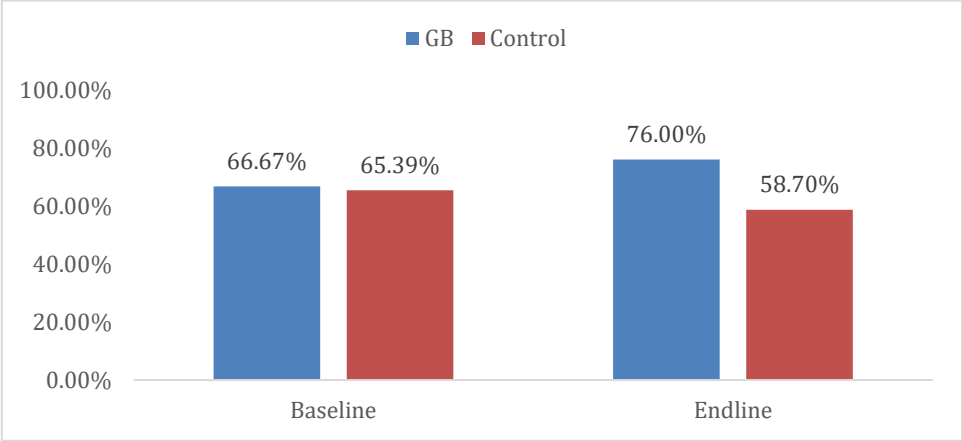


Figure 3. 27: Topics discussed in teachers' working group meetings (percentages)



We explored this issue further by comparing teachers' participation in the teachers' working group meetings and their satisfaction with these meetings. Teachers were given three options to express their level of satisfaction with the meetings: satisfied; adequate; and not satisfied. Figure 3.28 shows the percentage of teachers who were satisfied with the teachers' working group meetings. We noted a positive trend in the Guru BAIK group with the percentage of teachers who were satisfied increasing by almost 10 per cent while the percentage decreased in the control group. This led to a positive net change difference for the Guru BAIK group of 16 per cent. There are two possible interpretations for this finding. Firstly, the quality of meetings among the Guru BAIK group could have been superior, for example due to more lively discussion and more challenging topics. Or, secondly, the Guru BAIK teachers may be more aware of the importance of the teachers' working groups and so they value them more compared to those in the control group.

Figure 3. 28: Teachers satisfied with the teachers' working group meetings (percentages)



Teachers' knowledge, attitudes and skills

Previous studies on teachers' contributions to student learning in Indonesia suggest that teachers' subject-matter knowledge contributes more to the student learning outcome than teachers' formal qualifications, such as experience, employment status or formal degree obtained (see van Trotsenburg *et al.* 2015). As regards that evidence, this baseline study compared teachers' proficiency in numeracy and literacy materials at the primary school level.

Teachers' proficiency was assessed using literacy and numeracy tests that were originally designed for grade four primary school students. The Centre for Educational Assessment in the Indonesian Ministry of Education and Culture adapted the tests from the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). The format of the items included multiple choice, matching, short answers, drawing tasks and short essays. The high complexity items allowed the test takers to get half score if they respond with an only partially correct answer.

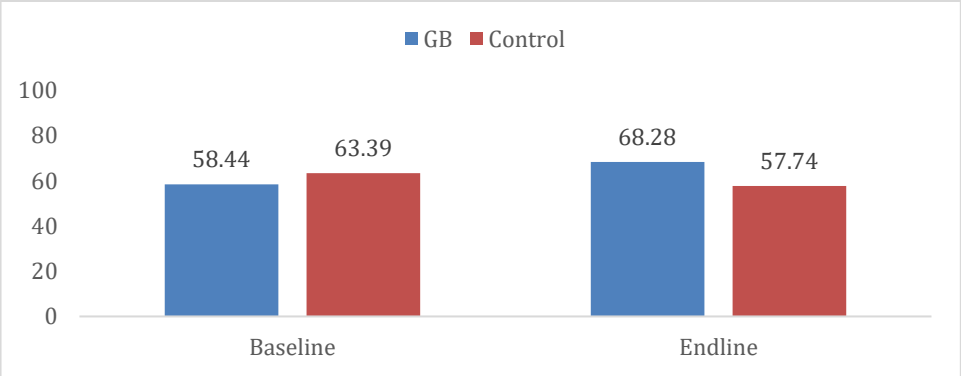
The Guru BAIK process did not include any sessions specifically aimed at improving teachers' subject-matter knowledge in literacy and numeracy although teachers were given various options to gain more knowledge (by reading books, peer discussions, browsing on internet or YouTube) that suited their needs. If teachers' subject matter knowledge has improved then it could be due to the peer discussions during the problem-solving process and the result of their own initiatives in looking for more teaching resources.

The literacy test assessed both lower and higher order thinking skills and was composed of the following cognitive process domains:

- Focusing on and retrieving explicitly-stated information (20 per cent);
- Making straightforward inferences (30 per cent);
- Interpreting and integrating ideas and information (30 per cent); and
- Examining and evaluating content, language and textual elements (20 per cent).

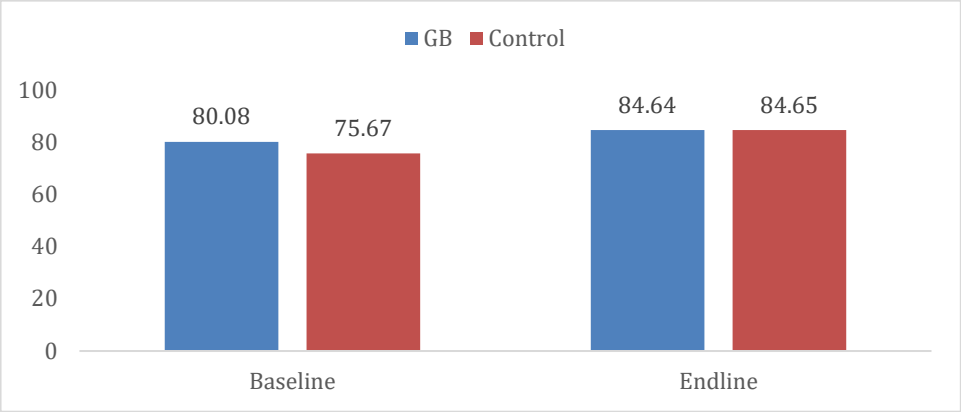
The results shown in figure 3.29 confirm an increase in teachers' average score by almost 10 among the Guru BAIK group. Meanwhile, in the control group, the average score decreased by 5.65. Combining these two findings, the net effect produced by the Guru BAIK intervention on teachers' average scores was more than 15 with statistically significant differences in score averages in the endline survey.

Figure 3. 29: Teachers' average literacy score (out of 100)



The numeracy assessment tools consisted of items related to: number (50 per cent); geometric shapes and measures (35 per cent); and data display (15 per cent). It assessed three different levels of thinking: knowing (40 per cent); applying (40 per cent); and reasoning (20 per cent). Figure 3.30 shows the change in teachers' average numeracy score from the baseline to endline surveys. The average numeracy score for the Guru BAIK group increased by 4.56, while the average score for the control group increased by 8.98. Consequently, the net change difference was negative for the Guru BAIK group although the difference in the endline scores was not statistically significant. Nevertheless, this finding needs further investigation.

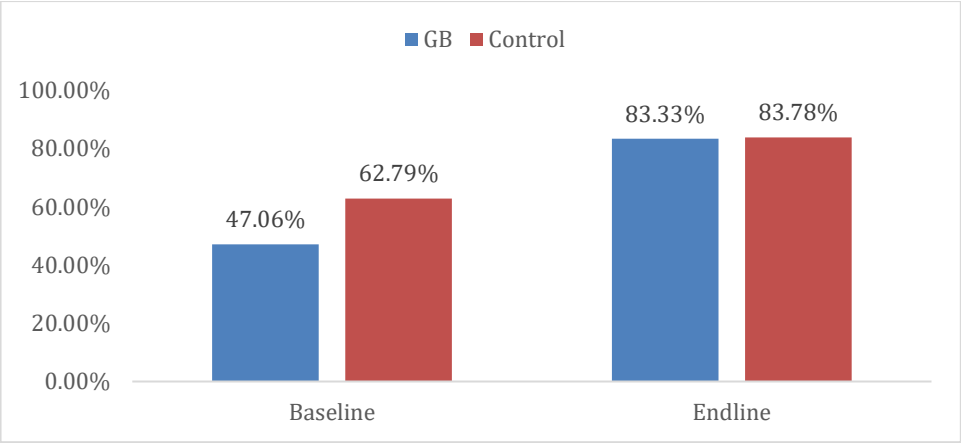
Figure 3. 30: Teachers' average numeracy scores (out of 100)



Turning to the internal factors that may contribute to teaching performance, an issue that emerges from research over the past decade is whether teachers have a growth mindset. Dweck (2008) found that a teacher's growth mindset had a large effect on students' achievement and learning progress. A growth mindset is a belief that intelligence or talent can be developed over time by learning. People with fixed mindsets believe that intelligence is a trait that people are born with and it cannot be changed (Dweck, 1996). Dweck (2008) believes that students are sensitive to how adults value their effort and intelligence and quickly pick up that belief and act accordingly.

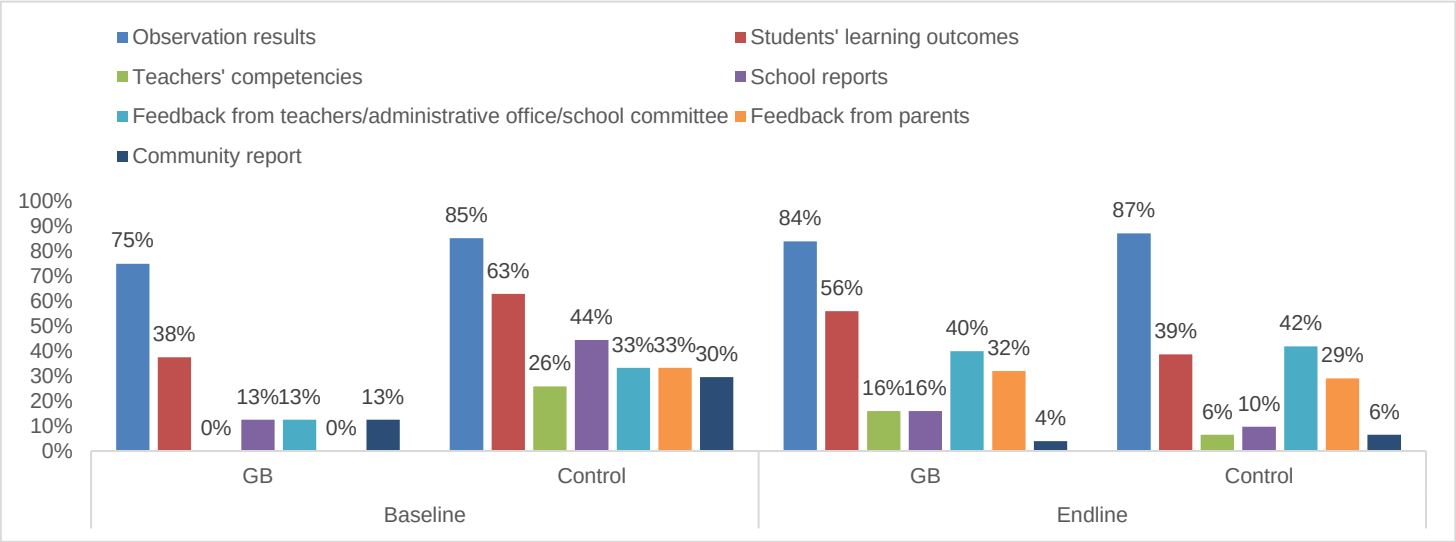
This pilot examines whether there was a shift in teachers' mindset in terms of using data to identify learning problems. The results presented in figure 3.31 show an increase in the proportion of teachers in the Guru BAIK group who identified learning problems based on evidence and by using data. There was a greater difference from the baseline to the endline survey in the Guru BAIK group than in the control group (36 per cent and 21 per cent respectively).

Figure 3. 31: Teachers who identify problems using data (percentages)



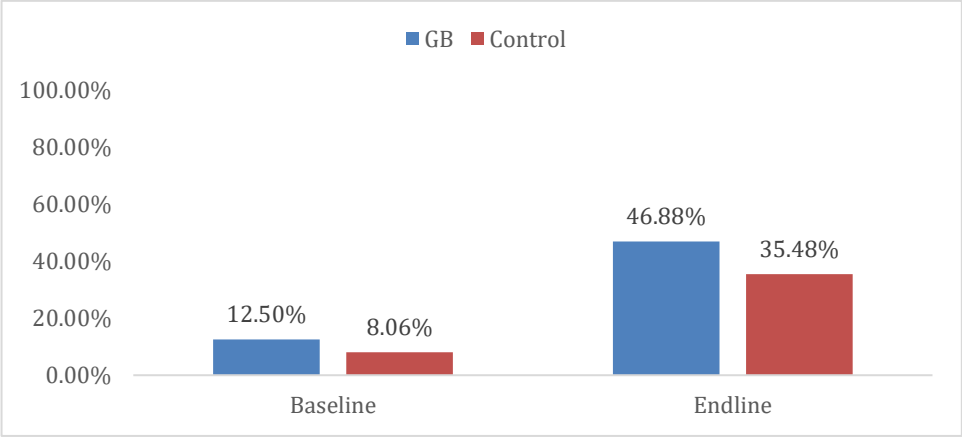
Teachers used various sources of data, ranging from students' learning outcomes to community reports. As illustrated by figure 3.32, observation results were used by most teachers in both the Guru BAIK and the control groups, according to baseline and endline data. The highest net change difference can be observed in the reliance on students' learning outcomes (almost 43 per cent).

Figure 3. 32: Types of data used to identify learning problems



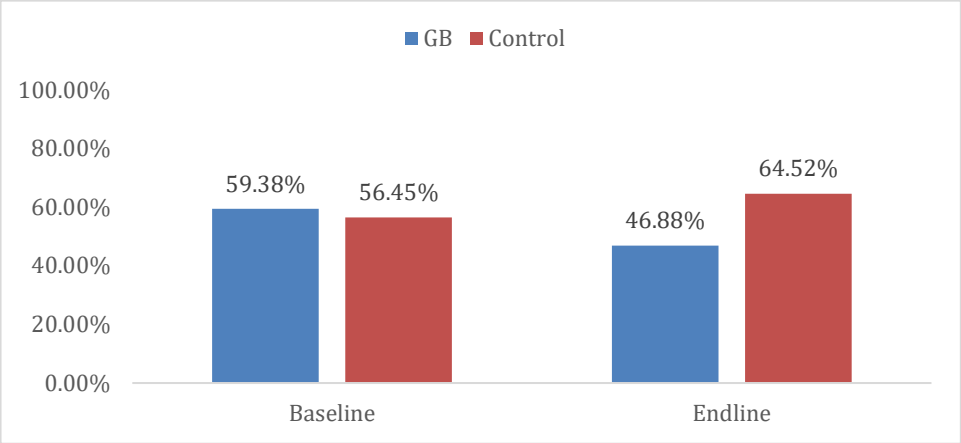
Another aspect of teachers' mindset shift that this pilot examined was their awareness or perception of the need to develop their teaching competencies. Teachers were asked whether they were aware that their knowledge and quality of teaching should be improved. Figure 3.33 shows that both in the baseline and the endline surveys, a higher proportion of teachers in the Guru BAIK group responded positively and the percentage increase was also significantly higher in this group relative to the control group.

Figure 3. 33: Teachers who were aware that their knowledge and quality of teaching should be improved (percentages)



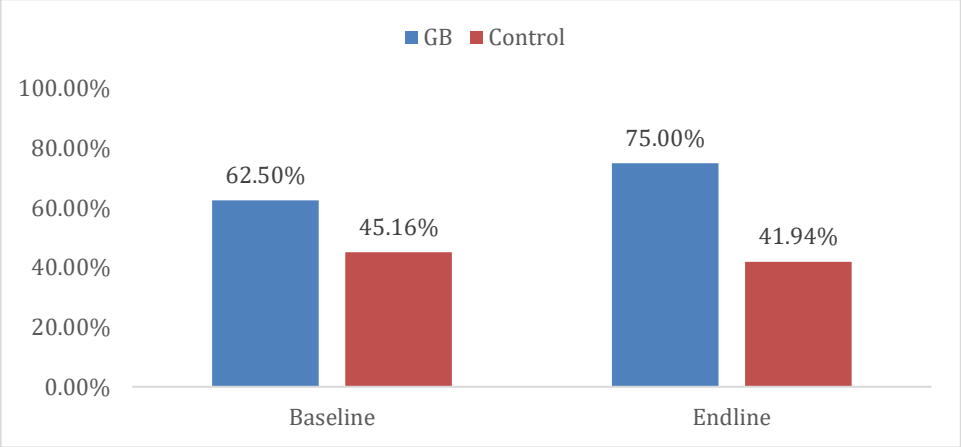
Teachers' attitudes were also captured by their absentee rates. Although in some cases teachers in the Guru BAIK group were absent due to training, the proportion of absent teachers was still lower in this group than in the control group in the endline, as shown in figure 3.34. This led to a negative net change difference of almost 21 per cent.

Figure 3. 34: Teachers who were absent from school (percentages)



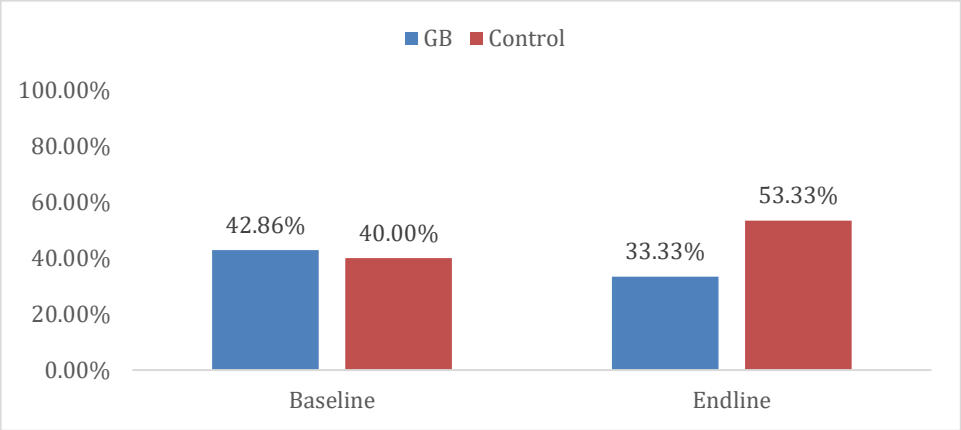
One important indicator in gauging improvements in teachers' skills is their participation in training relating to learning techniques and methods, materials, the curriculum and class action research. The study confirmed that the proportion of teachers being assigned to training courses among the Guru BAIK group rose by more than 12 per cent in the endline study. In contrast, in the control group, the percentage decreased by 3.22 per cent, leading to a net effect for Guru BAIK of almost 16 per cent (figure 3.35). Based on the data from the endline survey, differences in the proportion of teachers assigned to training in the Guru BAIK and control groups were statistically significant. This might relate to the shift in teachers' mindset making them more aware that they need to develop their teaching competencies.

Figure 3. 35: Teachers assigned to training (percentages)



Information on improvements in teachers' skills, knowledge and attitudes were also gathered from school principals, supervisors and parents. Firstly, school principals were asked whether they were satisfied with their teachers' knowledge and the results were compared over time between the Guru BAIK and control groups. Based on the results presented in figure 3.36, the proportion of school principals who were satisfied with teachers' knowledge among the Guru BAIK group went down from almost 43 per cent to 33 per cent while in the case of the control group the proportion increased by 13 per cent.

Figure 3. 36: School principals who were satisfied with teachers' knowledge (percentages)



A similar trend was found in the case of school principals' satisfaction with teachers' pedagogical competence. Figure 3.37 shows that the proportion of school principals who were satisfied decreased by almost 5 per cent among the Guru BAIK group and increased by 20 per cent among the control group.

Figure 3. 37: School principals who were satisfied with teachers' pedagogical competence (percentages)

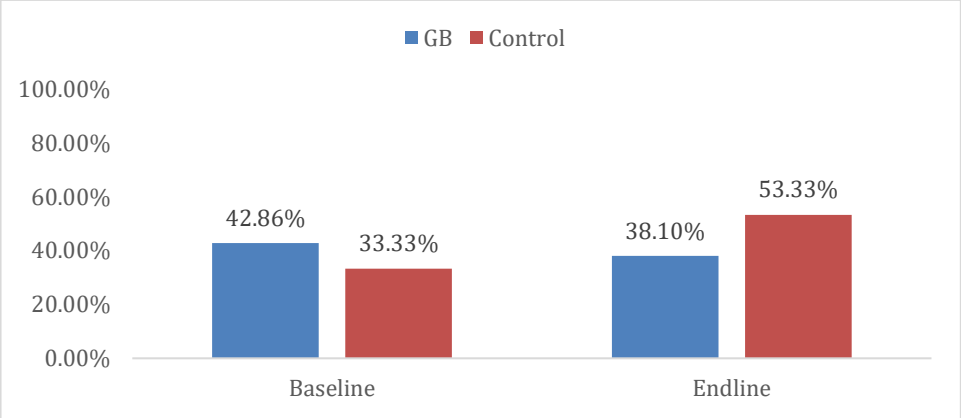
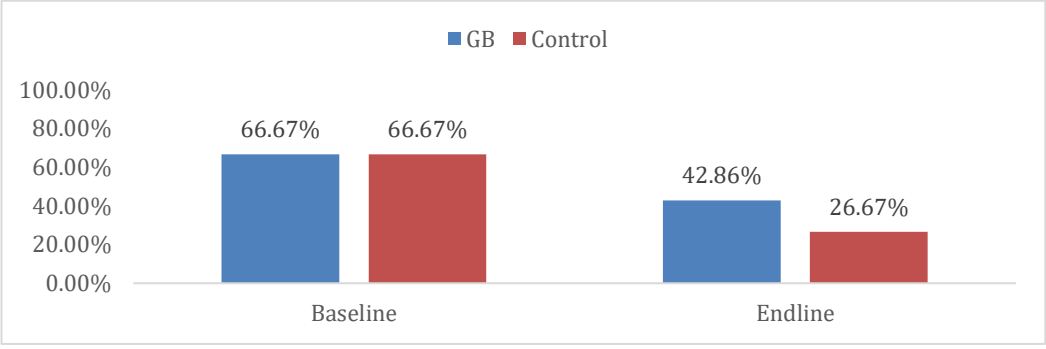


Figure 3.38 shows that in both the Guru BAIK and the control groups, school principals' overall satisfaction with teachers' understanding of the curriculum targets was lower in the endline survey. Nevertheless, a more significant decrease was documented in the case of the control group (40 per cent). This led to a net increase of more than 16 per cent for the Guru BAIK group which is statistically significant. This decrease might be due to the lack of involvement of school principals throughout the intervention period. They were only engaged at the beginning of the program (during socialisation) and at the end of the program.

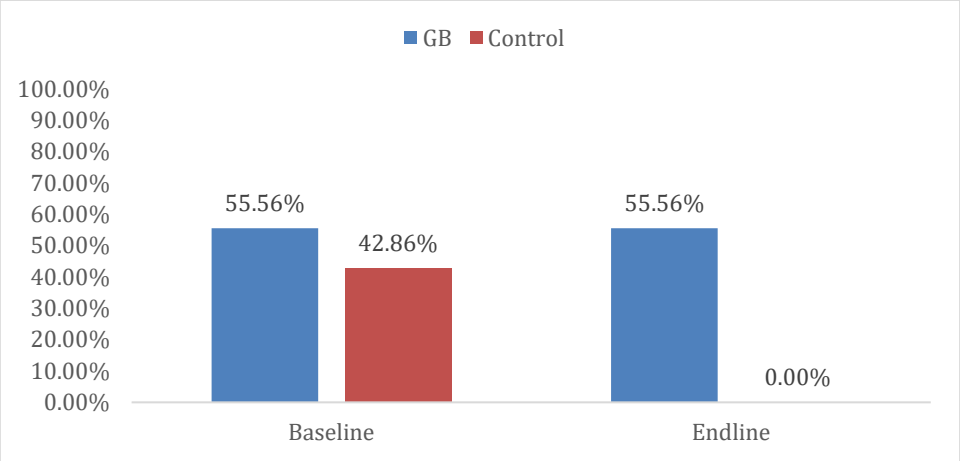
Figure 3. 38: School principals who are satisfied that teachers understand the curriculum's targets (percentages)



The lower satisfaction levels of school principals in the endline survey requires further investigation. Relating these findings to the result of the midline report, the low satisfaction levels among school principals in the Guru BAIK group in the endline survey might be the result of their lack of involvement in the Guru BAIK intervention.

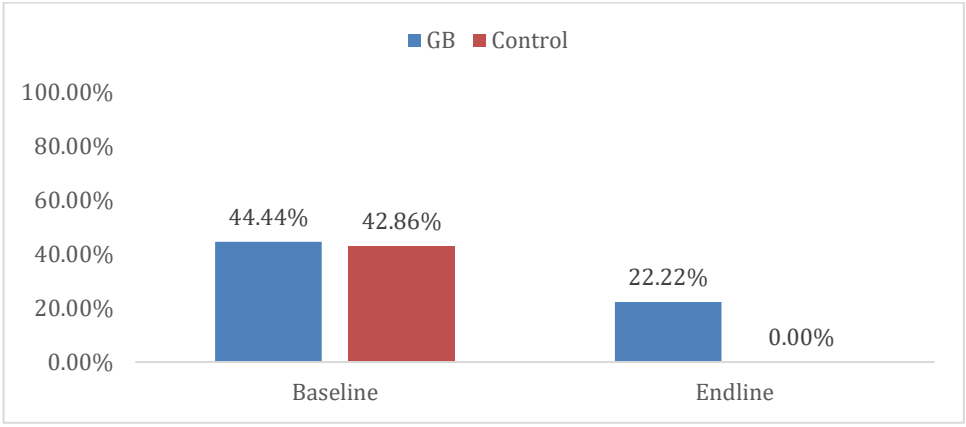
We asked school supervisors similar questions about their satisfaction with teachers' performance. We first asked supervisors whether they were satisfied with teachers' knowledge. While no change was documented in the Guru BAIK group (figure 3.39), in the control group none of the school supervisors was satisfied with teachers' knowledge in the endline survey.

Figure 3. 39: School supervisors who were satisfied with teachers' knowledge (percentages)



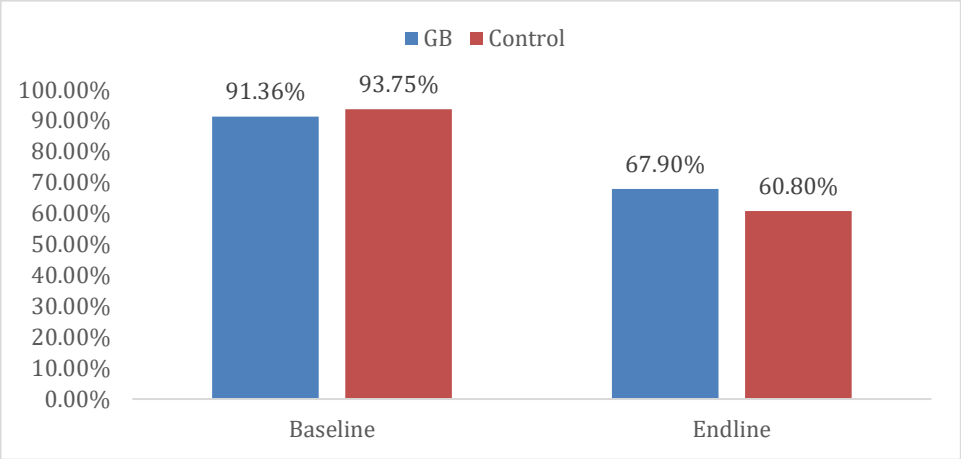
Supervisors tended to be less satisfied with teachers' pedagogical competence than with teachers' knowledge. Fewer supervisors in the Guru BAIK group were satisfied with teachers' pedagogical competence in the endline survey. However, similar to the previous question, none of the supervisors in the control group were satisfied with this matter in the endline survey (figure 3.40).

Figure 3. 40: School supervisors who were satisfied with teachers' pedagogical competence (percentages)



Parents were also asked about their satisfaction with the quality of the teachers and their behaviour and, as shown in figure 3.41, this decreased for both the Guru BAIK and the control groups (by more than 23 per cent and 33 per cent respectively). The Guru BAIK group recorded higher satisfaction levels in the endline survey (68 per cent versus 61 per cent for the control group) leading to a net positive effect for the Guru BAIK group of more than 9 per cent from the baseline figure.

Figure 3. 41: Parents who were satisfied with teachers' quality and behaviour (percentages)



3.2 CORRELATING FACTORS

This section presents the findings on factors that may correlate with students' performance in literacy and numeracy. Regressions were done on the results of the Guru BAIK baseline and endline surveys to shed some light on the factors that might contribute to the gaps in students' performance in West Nusa Tenggara. In our model, students' performances in literacy and numeracy are the dependent variables. Indicators included as independent variables comprised: students and parents' characteristics; teachers' qualifications and characteristics; and school and learning conditions. The results of the regressions are presented in table 3.5. However, as the number of sample is relatively limited for correlation analysis, the findings must be interpreted with caution.

Table 3. 5: Factors correlating with students' performance in literacy and numeracy

Independent variables	Baseline		Endline	
	Bahasa Coef (SE)	Math Coef (SE)	Bahasa Coef (SE)	Math Coef (SE)
class size	0.006*	0.001	-0.013	-0.002
	(0.003)	(0.004)	(0.009)	(0.009)
students' engagement	-0.172**	-0.241**	-0.047	0.463***
	(0.085)	(0.095)	(0.131)	(0.153)
use of active instruction	-0.128	-0.085	0.257	-0.218
	(0.103)	(0.119)	(0.185)	(0.188)
use of learning aides	0.061	-0.056	0.526**	-0.091
	(0.233)	(0.336)	(0.257)	(0.299)
school facility	0.248**	-0.119	0.91	-0.545
	(0.116)	(0.132)	(0.676)	(0.754)
attend PAUD/TK	-0.069	0.095	0.09	0.011
	(0.073)	(0.088)	(0.103)	(0.068)
father accompany	0.033	-0.017	-0.018	-0.023
	(0.031)	(0.033)	(0.042)	(0.037)
mother accompany	0.02	0.027	0.082*	0.015
	(0.03)	(0.034)	(0.041)	(0.061)
father's education (SMP or higher=1)	0.038	0.012	0.031	-0.019
	(0.034)	(0.036)	(0.036)	(0.043)
mother's education (SMP or higher=1)	0.038	-0.049	-0.028	-0.052
	(0.036)	(0.035)	(0.039)	(0.035)
natural logarithm of expenditure	0.021	0.017	0.047*	0.034
	(0.021)	(0.025)	(0.027)	(0.029)
principal supervision			0.492**	0.249
			(0.204)	(0.235)
student gender (female=1)	0.100***	0.056*	0.058	0.063*
	(0.028)	(0.031)	(0.042)	(0.036)
Raven test score	0.249**	0.278***	0.318***	0.144
	(0.099)	(0.098)	(0.115)	(0.138)
love to read	0.092*	0.015	-0.024	0.024
	(0.055)	(0.062)	(0.046)	(0.034)
seating position (front=1)	-0.02	-0.001	-0.066	-0.022
	(0.028)	(0.036)	(0.046)	(0.036)
student like Bahasa Indonesia	0.027		-0.026	
	(0.046)		(0.047)	
Bahasa Indonesia is easy	0.044		0.063	
	(0.037)		(0.043)	
student like Mathematics		0.022		0.073
		(0.038)		(0.048)
Mathematics is easy		0.012		-0.025
		(0.039)		(0.05)
excited going to school	0.290***	0.114	0.279***	0.463***
	(0.087)	(0.100)	(0.102)	(0.137)
gender of teacher (female=1)	-0.106	-0.025	0.064	0.116
	(0.102)	(0.097)	(0.073)	(0.103)
PNS	0.062	-0.096	0.085	0.148*
	(0.075)	(0.109)	(0.072)	(0.087)
attend training	-0.013	-0.001	-0.094	-0.07
	(0.055)	(0.067)	(0.094)	(0.085)
attend KKG meeting	-0.006	-0.001	-0.005	-0.007
	(0.004)	(0.004)	(0.006)	(0.007)
years of teaching	0.001	0.003	0.008	-0.014
	(0.004)	(0.005)	(0.009)	(0.009)
teacher growth mindset	-0.066	-0.155	0.275	0.281
	(0.109)	-0.096	(0.177)	(0.248)
teacher literacy test score	-0.096		0.058	
	(0.177)		(0.167)	
teacher numeracy test score		-0.700*		-0.221
		(0.358)		(0.136)
RPP Bahasa	0.08		0.096	
	(0.088)		(0.117)	
RPP math		0.04		0.1
		(0.091)		(0.134)
constant	-0.164	0.767	-1.581**	-0.352
	(0.358)	(0.593)	(0.591)	(0.717)
No. of Obs.	93	93	73	73
R-Squared	0.63	0.54	0.65	0.76

note: VCE robust
* p<0.10, ** p<0.05, *** p<0.010

Notes: Coef (SE) = standard error coefficient; PAUD = early childhood education; TK = kindergarten; SMP = junior secondary school; PNS = civil servants: KKG = teachers' working group; RPP = lesson plan; Obs = observations; R squared = coefficient of determination; VCE = variance–covariance estimate; p = calculated probability

Students and parents' characteristics

Among the indicators reflecting students' characteristics, gender, innate ability, enthusiasm for school and attitudes towards reading had significant correlations with students' performance in literacy and numeracy. With regard to gender, girls consistently outperformed boys except in the literacy scores in the endline survey. This is in line with a study by Burusic *et al.* (2012) in elementary schools in Croatia and another study by Duckworth and Seligman (2006). The most significant gender difference was seen in literacy performance during the baseline survey where, assuming other variables do not change in value,[12] it is predicted that girls' scores would be 0.1 higher than boys' scores.

Positive relationships between students' intelligence and their academic achievement is widely documented in the literature (see for example: Lopes-Soares *et al*. 2015; Costa and Faria 2018). In line with these findings, the results confirm that students with higher scores in innate ability tended to perform better in the endline survey, except in mathematics. Intelligence was shown to matter most in the case of literacy performance in the endline survey. A one point increase in the Raven score, the proxy of students' innate ability, is associated with a positive increase in literacy performance by more than 0.3.

Students' enthusiasm for school was shown to correlate significantly with performance, except in the case of mathematics in the baseline survey. Whereas in literacy performance there were no significant differences in terms of magnitude between the baseline and endline results. In numeracy, the difference in the nature of the relationships between enthusiasm and performance was remarkable. Based on the endline result, students who were excited about going to school can be expected to score higher in numeracy by 0.463. Gottfried (2018) also found this positive correlation between students' motivation and their academic achievement.

Students who love to read tended to score higher in literacy in the baseline survey. Fives (2015) found a similar conclusion among nine-year olds in Ireland. This is expected since attitudes to reading reflect an internal motivation for literacy learning.

In addition to students' internal characteristics, variables relating to parents and socio-economic status might correlate significantly with academic performance. Parental involvement has been shown to correlate positively with students' academic achievement. Topor *et al*. (2011) identified two possible mechanisms for parents' role in children's education: engaging with children in increasing their self-perception of cognitive competence and engaging with teachers and schools in promoting stronger student–teacher relationships. In this study, we examined whether students who had support from their fathers and mothers in studying performed differently from those who did not have support. The result showed that students who had support from their mothers tended to score higher in literacy. Meanwhile, no significant difference was found in the case of fathers' support.

Meanwhile, socio-economic status only made a difference in literacy in the endline survey with students from higher socio-economic levels tending to perform better.

Teachers' characteristics and teaching practices

The following indicators relating to the characteristics of teachers and their performance correlate with students' performance: students' engagement in learning; teachers' subject knowledge in mathematics (as reflected by numeracy scores); teachers' permanent status; and the use of learning media.

We identified a positive effect from the Guru BAIK intervention in the relationship between students' engagement and academic achievement. Research suggests a positive correlation between student engagement and academic performance (see, for example, Lee 2014) but this study found a negative correlation between students' engagement and both their literacy and numeracy performances, based on the regressions on the baseline data. This raises further questions about teaching practices as, based on

[12] Further analyses on the change of the dependent variable due to the change in one independent variable always assume that the values of the other independent variables are constant – *ceteris paribus* (all being equal).

observations, students in more engaging classes tended to score lower. However, regressions using the post-implementation data show that the correlation between students' engagement and their numeracy scores was positive and significant. It can be predicted that students in engaged classes tended to score higher by 0.463 points. Meanwhile, the significant negative relationship between engagement and literacy dissapeared by the endline survey.

An interesting result was seen in the negative relationship between teachers' numeracy scores and students' performance in mathematics in the baseline survey. This finding is counterintuitive since previous studies have shown the positive effect of teachers' subject knowledge on students' performance (for example, see Metzler and Woessmann 2012). However, such negative correlation was absent in the endline survey.

Teachers' permanent civil servant status (PNS) had a positive and significant correlation with students' mathematics performance in the endline survey. This indicates that teachers having a more secure employment status can have a positive effect on students' achievements. Meanwhile, there was a positive and significant relationship between the use of learning media and literacy scores in the endline study. One interpretation of this finding is that the Guru BAIK intervention helps Bahasa Indonesia teachers use learning aids more effectively. As Krolak (2005) argued, investing in learning aids like books and libraries is a critical factor for enhancing literacy skills and sustaining these skills for life.

School conditions

Factors reflecting the condition of schools have been shown to affect the quality of the learning process (Hasbullah *et al*. 2011) and some indicators reflecting this were included in this study. The result showed that the condition of schools generally tended to be less pertinent than students, parents and teachers' characteristics. Among the indicators examined in the regression models, school facilities had significant positive correlation with literacy in the baseline survey while the positive correlation between principals' supervision and literacy performance was only found in the endline study.

4. CONCLUSION

The key overall findings of the study for students were that students' literacy and numeracy scores in the Guru BAIK group increased compared to the control group. The highest effect of the intervention was seen among students in grade one with the effects gradually decreasing until grade three. There were different effects on grade four students which were positive for numeracy and negative for literacy. The Guru BAIK intervention had a higher effect on both literacy and numeracy compared to Gema Literasi. Different students' perceptions on Bahasa Indonesia and mathematics were documented and while there was a positive net change in students' attitudes to Bahasa Indonesia among the Guru BAIK group in the endline, a negative net change was recorded in their attitudes to mathematics. Another inconsistent trend was the negative net change in enthusiasm for school among students in the Guru BAIK group.

Positive findings for teachers were the possible effect of Guru BAIK on indicators relating to teaching practices. These were reflected in the following: engaging classes; using active instruction; using learning media; giving assignments to enhance students' literacy; drawing up workplans; and attending teachers' working group meetings. For subject knowledge, teachers' literacy scores showed a positive net change while the numeracy scores showed a net negative change. The survey also documented positive mindset shifts among teachers with the proportion of teachers using data, especially regarding students' learning outcomes, to identify problems also showing a positive net change.

Although positive improvements were found in most indicators relating to students and teachers in the Guru BAIK group, the perspectives of school principals were not entirely aligned. Among those in the Guru BAIK group, levels of satisfaction with teachers' performance in learning materials, paedagogical skills and students' learning outcomes showed a negative net change difference. On the other hand, supervisors' levels of

satisfaction on the same aspects tended to increase. Parents' perceptions of the benefits of schools for improving literacy and numeracy showed a negative net change difference for the Guru BAIK group although positive net change differences were seen in other aspects.

In conclusion, the Guru BAIK intervention led to improvements in most indicators relating to teachers and students. However, these were not entirely evident in school principals' perspectives. One possible explanation is that school principals were not fully informed about these achievements because of their lack of involvement. Some results must be taken with caveats, considering that the endline and baseline surveys were conducted at different times in the academic year and that the endline survey was affected by the earthquakes in West Nusa Tenggara.

REFERENCES

Burusic J, T Babarovic and M Serie (2012) 'Differences in elementary school achievement between girls and boys: Does the teacher gender play a role?', *European Journal of Psychology of Education* 27(4): 523-538.

Bruns, Barbara; Luque, Javier; Bruns, Barbara ; Luque, Javier. 2014. Great teachers : how to raise student learning in Latin America and the Caribbean (English). Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/432391468089383429/Great-teachers-how-to-raise-student-learning-in-Latin-America-and-the-Caribbean

Costa A and F Faria (2018) 'Implicit theories of intelligence and academic achievement: a meta-analytic review', *Frontiers in Psychology* 9:829, doi: 10.3389/fpsyg.2018.00829.

Dorans, Neil J, Moses Tim P and Eignor Daniel R (2010) 'Principles and Practices of Test Score Equating'. Retrieved from https://www.ets.org/Media/Research/pdf/RR-10-29.pdf

Duckworth A and M Seligman (2006) 'Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores', *Journal of Educational Psychology* 98:198–208.

Dweck CS (1996) Implicit theories as organizers of goals and behavior". In The psychology of action: Linking cognition and motivation to behavior, edited by PM Gollwitzer and JA. Bargh. pages 69-90. New York: Guilford Press.

Dweck CS (2008) Mindsets and math/science achievement. Princeton: Institute for Advanced Study. Retrieved from http://www.growthmindsetmaths.com/uploads/2/3/7/7/23776169/mindset_and_math_science_achievement_-_nov_2013.pdf.

Fives A (2016) 'The association of attitude to reading and reading achievement among a representative sample of nine year olds in Ireland', *Reading Psychology* 37(1):27–54.

Gottfried AE (2018) 'Academic intrinsic motivation: theory, assessment and longitudinal research', in AJ Elliot (ed) *Advances in motivation science volume 6*, Cambridge USA and London: Academic Press.

Hasbullah A, WZ Wan Yusoff, M Ismail and P Vitasari (2011) 'A framework study of school facilities performance in public primary school of Batubara district in Indonesia', *Procedia – Social and Behavioral Sciences* 15: 3708-3712.

Krolak L (2005) 'The role of libraries in the creation of literate environments', Paper commissioned for the EFA Global Monitoring Report 2006, Literacy for Life.

Lee JS (2014) 'The relationship between student engagement and academic performance: Is it a myth or reality?', *Journal of Educational Research* 107:3.

Lopes-Soares D, GC Lemos, R Primi, LS Almeida (2015) 'The relationship between intelligence and academic achievement throughout middle school: the role of students' prior academic performance', *Learning and Individual Differences* 41: 73–78.

Metzler J and L Woessmann (2012) 'The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation', *Journal of Development Economics* 99(2):486–96.

Topor DR, SP Keane, TL Shelton and SD Calkins (2011) 'Parent involvement and student academic performance: a multiple mediational analysis', *Journal of prevention & intervention in the community* 38(3): 183–197.

van Trotsenburg A, RA Chaves, C Costin, H Patinos, A Ragatz (2015) *Teacher certification and beyond An empirical evaluation of the teacher certification program and education quality improvements in Indonesia*. Jakarta: Education Global Practice, World Bank.

Widhiarso W (2018) 'Applying generalized analysis of covariance for accommodating individual differences: a study effect of faking on personality test', *Anuario de Psicología* (2018) 48:81–91.