

Strengthening Teachers Accountability to Reach all Students (STARS)

World Bank SIEF

Milestone 2: Baseline Report

November 2018

Anne Fitzpatrick, Assistant Professor of Economics, College of Liberal Arts,
University of Massachusetts Boston

Adrienne Lucas, Associate Professor of Economics, Lerner College of
Business and Economics, *University of Delaware*

Sabrin Beg, Assistant Professor of Economics, Lerner College of Business
and Economics, *University of Delaware*

Renaud Comba, Henry Atimone, Bridget Konadu Gyamfi, Joyce Jumpah,
and Edward Tsinigo, *Innovations for Poverty Action*

Executive Summary

Previous research in Ghana and India demonstrated the effectiveness of "targeted instruction"—teaching students at their level of knowledge, not their grade level. The Teacher Community Assistant Initiative (TCAI) in Ghana found that this model increased learning by students despite limited teacher take-up of the program. Low adherence to this intervention, and interventions more broadly, raises challenges as governments attempt to scale-up effective interventions. One potential factor impacting low adoption rates among teachers is the lack of managerial support, which may prevent teachers from thoroughly implementing the program. Through a randomized controlled trial, this project will (1) test the effect of training teachers on targeted instruction and (2) test whether additional management training of HT (head teachers, i.e. school principals) and CSs (circuit supervisors, i.e. middle-level management responsible for a subset of schools within a district) increases the quality of implementation of targeted instruction and student outcomes. This study works within the system to improve educational outcomes. Ghana Education Services (GES), National Council for Curriculum and Assessment (NaCCA), and the National Inspectorate Board (NIB) have designed the materials and trained the teachers, HT, and CSs.

As part of evaluation activities prior to the training, Innovation for Poverty Action (IPA) enumeration teams visited 211 schools, interviewing 210 HT and 671 P4 through P6 teachers and interviewing and testing 5,893 P4 and P5 pupils in both English and math. This report summarizes the baseline findings. Across the treatment groups, the averages of the measures that we collected are statistically similar. Adult respondents report less work burn-out and stress than focus group discussions suggested. Nothing in the baseline data collection or analysis portends issues in the research design or future data collection.

Content

Executive Summary	2
Content	3
List of Tables	5
List of Figures	6
Abbreviations	7
Section 1: Introduction and Background of the STARS Project	8
1.1 Background	8
1.2 The STARS Project	9
1.3 Description of the STARS Intervention	10
1.3.1 Targeted instruction	10
1.3.2 Management training	11
1.4 Implementing Partners and Stakeholders	11
Section 2: Evaluation Design	13
2.1 Theory of Change	13
2.2 Study Design	14
2.3 Evaluation Scope	15
2.4 Evaluation Questions	15
2.5 Sample and Sampling Procedure	16
2.6 Outcome Measures	17
2.7 Baseline Data Collection	18
2.7.1 Baseline questionnaires and content	18
2.7.2 Field staff recruitment and training	20
2.7.3 Baseline coverage	20
2.7.4 Data collection challenges	21
Section 3: Baseline Findings and Validation Tests	22
3.1 Description of Data Characteristics	22
3.1.1 Circuit Supervisors	22
3.1.2 Head Teachers	23
3.1.3 Teachers	23
3.1.4 Students	24
3.2 Baseline Equivalence on Outcomes Measures	25

3.3	Risks to Impact Evaluation Design	26
3.4	Analysis and Lessons Learned from Baseline	27
Section 4: Intervention Implementation		28
4.1	Implementation of STARS Intervention	28
4.2	Intervention Piloting.....	29
4.2.1	Observations from the pilot.....	30
4.3	Key Lessons from Implementation	31
Section 5: Conclusions and Next Steps		32
5.1	Conclusions	32
5.2	Next Steps	32
Appendix		33
References		37

CONFIDENTIAL

List of Tables

Table 1: Baseline Coverage Rate by Instrument	21
Table 2: CS characteristics, by Treatment Group.....	33
Table 3: HT characteristics, by Treatment Group.....	34
Table 4: Teacher characteristics, by Treatment Group.....	35
Table 5: Child characteristics, by Treatment Group.....	36
Table 6. Timeline for Key Implementation Activities	29

CONFIDENTIAL

List of Figures

Figure 1: The Study Design.....	15
Figure 2: Distribution of Combined Student Latent Ability by Grade	25

CONFIDENTIAL

Abbreviations

CS	CS
DDE	District Director Education
DDSup	District Director Supervisor
DTST	District Teacher Support Team
GES	Ghana Education Service
HT	HT
INSET	In-Service Education and Training
IPA	Innovations for Poverty Action
MoE	Ministry of Education
mSRC	Mobile School Report Card
NaCCA	National Council for Curriculum and Assessment
NIB	National Inspectorate Board
NTC	National Teaching Council
P4	Primary class 4
P5	Primary class 5
P6	Primary class 6
STARS	Strengthening Teachers Accountability to Reach all Students
TaRL	Teaching at the Right Level
TCAI	Teacher Community Assistant Initiative
TI	Targeted Instruction

Section 1: Introduction and Background of the STARS Project

1.1 Background

Many Sub-Saharan African countries have made considerable progress towards the achievement of the Goal 4 of the United Nations Sustainable Development Goals that encourages countries to “Ensure inclusive and quality education for all”. Evidence of this progress is the increasing investments in promoting access to primary education. As in many countries in sub-Saharan Africa, primary enrollment rates in Ghana have increased substantially in recent years, but students’ learning levels have not matched this progress. For example, the United National International Children’s Emergency Fund’s (UNICEF) Global Initiative on Out of School Study (2012) report indicated that 89% of Ghanaian children were enrolled in primary school, but only 8% met the academic standards for their grade. Notably, available evidence suggests an astounding number of students performing far below competency levels and without basic literacy and numeracy skills (Pritchett 2013; UWEZO, 2014). Similarly, the 2016 National Education Assessment found that fewer than 25% of pupils in primaries 4 and 6 were considered competent in mathematics and only about 37% were competent in English. The nationwide aggregates mask important differences across regions: Northern, Upper East, and Upper West regions had the lowest average scores.

These gaps in learning outcomes among primary students highlight heterogeneity in primary school classrooms. Strict adherence to the official level of the curriculum causes some students to fall further behind, with learning gaps growing each year. Therefore, while grade-specific learning standards are important, teaching to children at their own level in order to decrease the gap between their own knowledge and the official standards is also crucial and offers an inclusive approach to learning.

One promising way to enhance teaching quality and learning outcome in developing countries is *targeted instruction* - teaching at the knowledge level of the student instead of grade level. Previous research in Ghana, India, and Kenya demonstrated that targeted instruction is an effective and cost-effective strategy to increase student learning (Banerjee et al., 2007; Duflo et al., 2011; Duflo et al., 2018). The study in Ghana, known as the Teacher Community Assistance Initiative (TCAI) is particularly relevant to the present study. TCAI revealed that targeted instruction successfully improved learning outcomes, although only a fraction of teachers implemented the program. These results suggest that with weak accountability, monitoring, and supervisory support for teacher-led programs, governments will be challenged to achieve the same results as they move to scale. To scale up targeted instruction, governments will face the challenge of ensuring that all teachers implement the program as designed. Thus, research and

reflection must propose effective means to support teachers and boost their compliance with this pedagogical approach at scale.

Efforts to understand low compliance by teachers in the TCAI implementation revealed that lack of teacher motivation and absenteeism were contributing factors. On average teachers were absent 30% of the time, and even when present in the school, time-on-task was low. Only 15% of teachers trained in the method correctly implemented it during unannounced spot-checks (Duflo et al., 2018). This calls for concerted efforts to establish strong accountability, monitoring, and supervisory support to enhance broad take-up of the targeted instruction program.

Recent research has shown the importance of school management in explaining variation in student learning outcomes. Providing additional management training and similar “expert consulting” services can improve managerial capacity, increasing productivity and organizational functioning in the developing country context (Bloom et al., 2013; Mano et al., 2012). In South Africa, Cilliers et al. (2018) showed that improved management support and feedback (“coaching” teachers) was both more effective at improving student test scores than traditional pre-term teacher training programs. Teacher coaching was also more cost effective. Notably, while training alone had an impact of 0.12 standard deviations, including coaching increased student learning by 0.24 standard deviations (Cilliers et al., 2018).

Improving the learning outcomes of students through an efficient school management system is a top priority of many governments and policymakers such as the Ministry of Education and the Ghana Education Service. UNICEF has been supporting MoE / GES in this effort. Given the persistent low learning levels demonstrated in repeated reports on learning in Ghanaian primary schools, embedding teacher-led targeted instruction with improved school management practices could be a way to address persistent low learning achievement. It may also provide useful evidence on ways in which to scale up a successful model of providing remedial education to pupils who lag behind.

1.2 The STARS Project

Focus groups with teachers, Head Teachers (HT) and Circuit Supervisors (CS) following TCAI identified that weak management policies were a contributing factor to low take-up of targeted instruction. Therefore, STARS was developed to replicate and improve upon the success of TCAI by including a management arm. STARS will encourage fidelity to implement targeted instruction by leveraging the role of existing education personnel, teacher managers. STARS will build the capacity of HT and CS by improving their coaching and mentoring skills. As a result, these supervisors will support teachers with monitoring resources to implement the targeted instruction pedagogy with fidelity. The project seeks to explore (a) how teachers can be motivated and empowered to implement new approaches that have the potential to improve learning levels; and

(b) how supervision can be strengthened to provide coaching and monitoring support for teachers to implement new pedagogical approaches.

1.3 Description of the STARS Intervention

The main interventions under the STARS project are targeted instruction and enhanced management.

1.3.1 Targeted instruction

Targeted instruction under the STARS project focuses on equipping the classroom teachers from primary 4 to primary 6 to teach at the knowledge level of the child. The focus will be to help pupils in the targeted grades who are performing below grade level in English and Math acquire the needed fundamental skills to enable them to perform at grade level. The project will build the capacity of teachers, HT, and CS on how to deliver the targeted instruction (TI) pedagogy. At the beginning of the school year, all pupils in primary 4 to primary 6 will be tested in Maths and English [using basic ASER-type tool] and grouped in levels. For one hour a day, 4 days a week and 8 weeks a term for one academic year, teachers in randomly selected schools across 20 districts will deliver TI to pupils based on the learning levels. Teachers will use formative assessments to track the progress of pupils and re-test pupils at the end of the term. Children will be tested as follows a) at the beginning of Term 1; b) beginning of Term 2; c) beginning of Term 3 and d) end of Term 3. The various assessment points provide an opportunity for the teacher to track progress and move children who make progress into higher levels. The assessment seeks to achieve two goals:

- a. *Learning goal for literacy:* At least 70% of children should be able to read a story in English with comprehension by the end of one academic year;
- b. *Learning goal for math:* All children should be able to recognize numbers up to 100 with the place value, and 70% of Children should be able to do basic operations including the division of two digits by 1 digit by the end of one academic year.

During the TI hour, teachers will engage children in whole class activities, group, and individual activities.

- a. *Class activities:* The whole classes activities are designed in such way that all children should understand how to be engaged in different activities, come out of their inhibition, feel confident to converse freely in peer groups and with the teacher/facilitator;
- b. *Group activities:* Since children are divided according to their levels, they can be engaged with the activities meant for them. For example, the 'beginner and letter' level children are

helped to work in groups to recite ‘syllable chart’, ‘recognize letters/syllables from the chart’ and ‘create words – known/unknown’ from syllable chart;

- c. *Individual activities:* Once children get used to performing certain activities such as creating ‘mind-map’ and ‘reading a simple paragraph’, they can work independently to practice reading and writing.

1.3.2 Management training

The CS and HT will receive additional training on the best practices to mentor and support teachers as they implement the program. CS and HT will receive a succinct resource manual on how to perform specific coaching and mentoring support activities to teachers. In addition, CS and HT will receive positive reinforcement SMS text messages directly from GES that will include tips on how to be a better mentor/coach and perform effective monitoring and classroom observations. The CS and HT will have access to a GES-run Help Desk where teachers, CS or HT can call or text to receive prompt feedback on any issues that they are facing. During refresher training, CS and HT will undergo peer-learning activities within their respective enhanced management refresher training to share best practices and lessons learned with other CS and HT.

1.4 Implementing Partners and Stakeholders

The STARS project is a partnership among many institutions - Ministry of Education (MoE) [National Inspectorate Board (NIB), National Teaching Council (NTC), and National Council for Curriculum and Assessment (NaCCA)], Ghana Education Service (GES), United National International Children’s Emergency Fund (UNICEF), and Innovations for Poverty Action (IPA). Under this partnership, MoE/GES will leverage on its existing structures to implement the project in selected schools. UNICEF will provide technical and financial support, whereas IPA will provide technical support in the intervention design and support the principal investigators to conduct an evaluation of the project. GES as the implementing organization on the STARS project will lead the implementation of the project by setting up structures and systems for successful implementation.

- a. *The National Council for Curriculum and Assessment.* The NaCCA leads the development of TI materials. NaCCA will constitute a Resource Development Team (RDT) made up of experts from the education sector to a) conduct an initial review of existing GES and TCAI materials and resources, as well as b) design and develop appropriate materials relevant for the STARS intervention from primary 4 to primary 6;
- b. *The National Inspectorate Board.* The NIB, a division under MoE, is responsible for the development of the CS and HT intervention materials (manuals, content for text message reminders, etc.). NIB will manage logistics and oversee the manual development process,

provide status updates to stakeholders, and submit final copies of manuals to the STARS Core Team ahead of project implementation;

- c. *The National Teaching Council.* The NTC, a division under MoE, is responsible for the development of training materials and training of teachers on the TI pedagogy. As part of its role, NTC will constitute a team of Core trainers who will be responsible for (a) developing training guides and facilitator manuals and (b) conduct a train the trainer session for district training teams. NTC will supervise the base intervention training in all districts and manage all the logistics associated with carrying out these tasks. NTC will be required to provide status updates to stakeholders and submit reports after each milestone activity to the STARS Core Team;
- d. *The Ghana Education Service.* The Basic Education Division (BED) of GES will support NaCCA, NIB, and NTC to develop relevant materials needed for the successful implementation of the STARS project. Specifically, the BED will support the development of the TI materials, CS, and HT Quick Reference Guides. The BED will participate in all partner engagement meetings including the quarterly core group meetings. The responsibilities of GES are outlined below.
 - a. Preparatory phase: GES is responsible for:
 - i. Organizing district orientation meetings and engagements to prepare the 20 UNICEF districts for the implementation of the project;
 - ii. Leading the implementation of the pilot phase in two districts; and
 - iii. Providing reports on the pilot to feed into the revision of the intervention design and coordinating the organization of the teacher-training program.
 - b. Implementation phase: GES is responsible for:
 - i. Facilitating the distribution of an official GES letter to all intervention schools;
 - ii. Coordinating with NTC to deliver the CS and HT Intervention training;
 - iii. Exploring the use of Mobile School Report Card (mSRC) to set up a text message platform;
 - iv. Hosting and managing a Help-Desk designed to answer/respond to CSs and HT questions on the implementation of the targeted instructing implementation; and
 - v. Providing status updates to all relevant stakeholders in a timely manner.

Section 2: Evaluation Design

2.1 Theory of Change

Students are in school, but not learning, especially those from marginalized groups. Moreover, students are at different levels of preparation, particularly first generation learners. Therefore, they do not benefit from instruction that is beyond their existing learning level. As a result of this heterogeneity within the classroom, teachers are unable to effectively teach the grade-level material to all students.

The TI intervention aims to increase learning levels of all students and will ensure learners of all genders and backgrounds have the same opportunities. TI divides students in grades 4 through 6 by their learning levels instead of their grade level. Our theory of change is that grouping students according to their learning level reduces within-class heterogeneity and allows teachers to effectively teach the relevant material. During the three TI hours, students who are behind receive instruction at their level, remedying basic skills and allowing them to catch-up to grade-level work. Teachers receive training in this method of teaching and a suite of teaching and learning materials, allowing them to focus on improving student learning given their existing skills instead of trying to finish the syllabus.

While TI is demonstrated to be effective, previous work found that inadequate managerial capacity and teaching support for TI were barriers for teachers to implement this pedagogy. This lack of managerial support limits the success of introducing TI at scale. The management intervention aims to increase the likelihood that teachers implement TI. The direct managers of teachers, HT and CSs, are often occupied with mundane paperwork and record-keeping instead of effective school leadership and mentorship. A teacher-led program is unlikely to be implemented thoroughly when supervisory personnel are detached from teachers and their teaching. The enhanced management program will fill this gap. Managers will be included in the TI training and offered additional training on managerial best practices—how to mentor, support, and observe teachers, and partner with them as a cohesive unit to be pedagogical leaders and improve student outcomes. HT and CSs compile data that is seldom used. This training will impart them with the tools and skills to effectively use data in decision-making—tracking teacher absenteeism and student learning and knowledge progression. Other topics of training include time management, performance reviews, and effective communication.

Our primary hypothesis is that management and leadership capacity is not a fixed trait of someone, but rather a characteristic that can be improved through education and training. We hypothesize that training more effective managers will get teachers into the classroom and improve their student engagement. We believe that ultimately this program will improve student learning, critical thinking skills, and prepare the next generation of Ghanaian children for a modern

economy. We expect it to succeed because parts of our complete program have already found positive results, albeit in other settings. A recent study in Indian garment workers recently showed that soft skills are meaningfully improved with specific, effective training (Adhvaryu et al., 2018). Help desks, or “hotlines” have been successful in settings ranging from Ugandan health programs to rural U.S. schools; SMS messages have been effective nudges in a large number of settings, ranging from appointment to loan repayment reminders. Moreover, this program is “demand-driven”. For example, the Help Desk and SMS message components were added after focus groups with HT and CSs on barriers to effective teaching practice. We believe that the additional managerial support will be key for ensuring teachers implement TI, teach effectively, and that ultimately student learning will improve.

2.2 Study Design

We will test STARS by implementing a randomized controlled trial in 140 circuits (groups of schools) within the 20 districts that UNICEF supports. The STARS study has two layers of randomization: circuit-level and school-level, resulting in three study arms (two treatment groups and one control group). Specifically, we randomly selected circuits to receive management training and randomly selected schools to receive training on targeted. The study design is pictured in Figure 1.

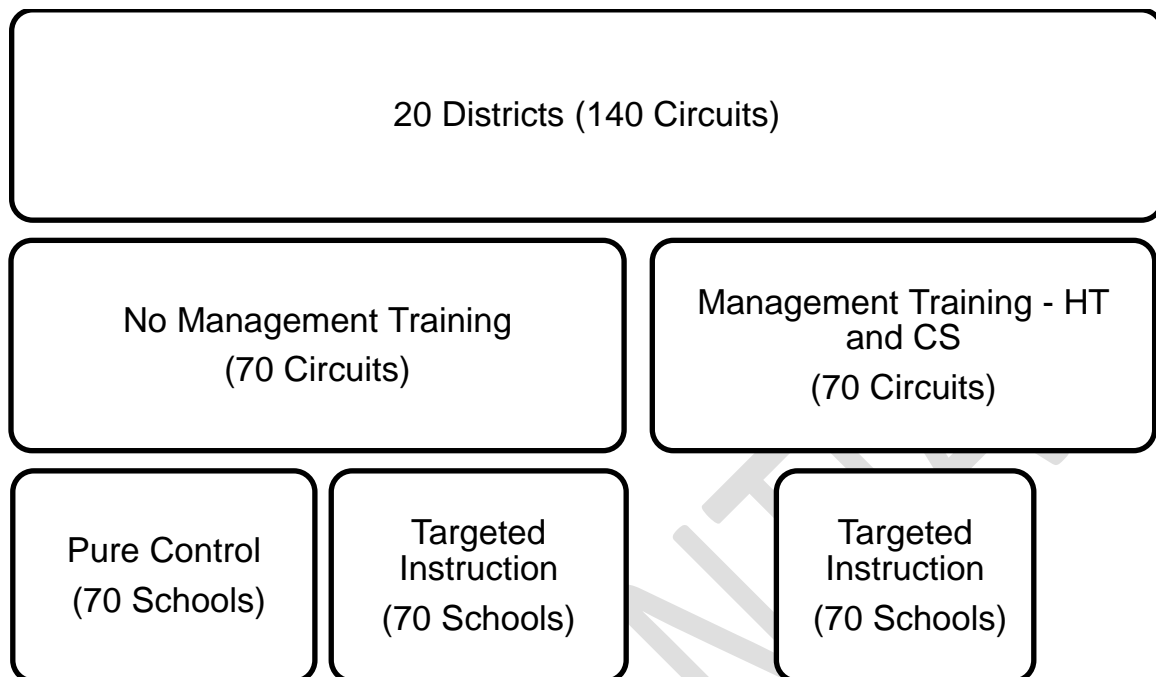
First, the circuits were randomly divided into two groups:

Group 1 (70 CSs) no management training: CSs in these circuits only receive the basic in-service teacher focused training on targeted instruction. Each Group 1 circuit has two study schools: one school is “pure control”, in which teachers receive no training on targeted instruction and the other school receives “targeted instruction” treatment, in which HT and teachers receive teacher focused training on targeted instruction. Assignment of “pure control” and “targeted instruction” at school level is random.

Group 2 (70 CSs) management training: CSs in these circuits receive the same teacher focused training on targeted instruction as Group 1 plus additional management training and support. Each Group 2 circuit has one study school. Within the study schools in Group 2, HT receives additional management training and resources and teachers and HT receive teacher focused training on targeted instruction (“targeted instruction”).

This study design is similar to a fully cross-randomized design, although we do not have any schools that receive the management intervention without targeted instruction. This design has the advantage of allowing us to measure the impact of TI alone compared to business-as-usual, and the impact of TI+Management compared to business-as-usual. It also allows us to compare the differential effect of the management intervention with TI-only, establishing effectiveness but also cost-effectiveness.

Figure 1. The Study Design



2.3 Evaluation Scope

The STARS project is being implemented in seven administrative regions (excluding Brong-Ahafo, Ashanti and Western Regions) of Ghana. STARS covers 20 UNICEF districts, 140 circuits, and 210 schools across these regions. To analyze the effects of the program, we will conduct four rounds of data collection – baseline, two midline observational studies, and a follow-up study. We completed a baseline survey in May/June 2018. We will also conduct two midline observational studies of classroom behavior [in October 2018 and February 2019] and a follow-up study in May 2019.

2.4 Evaluation Questions

The STARS project leverages existing personnel and structures in the Ghana education sector to improve student achievement and answer the following questions:

1. Does targeted instruction improve student test scores in upper primary grades?
2. Can monitoring, managing, and coaching performed by existing CSs and HT increase the likelihood that teachers implement targeted instruction?
3. Do the enhanced monitoring, managing, and coaching lead to higher learning gains?

4. What is the cost-effectiveness of targeted instruction alone as compared to targeted instruction with enhanced management training?

2.5 Sample and Sampling Procedure

Participants were sampled using a broad three-stage sampling design comprising the selection of the districts, circuits, schools, and individual participants. The various stages involved in selecting the participants are discussed as follow.

1. The sample universe is all 145 circuits in 20 UNICEF districts, We randomly assigned each study circuit to a treatment status, using district as a stratification cell. We combined two different administrative data sources to create the universe of schools and created a randomly ordered list of all target schools within each circuit. We then proceeded to contact all schools in the order in which they appeared on the list and administered the screening questionnaire based upon specific eligibility criteria;
2. To be eligible for the STARS project, a school must have non-zero enrollment in primary 4, primary 5, and primary 6 classes¹, non-shift² in primary 4-primary 6, and non-multi-grade³ in primary 4 through primary 6;
3. Some circuits are small and were anticipated to have few schools meeting the criteria. While Management circuits only required one school to fit the screening criteria, Non-Management circuits required two schools to fit the screening criteria. In order to ensure that circuit-level characteristics were balanced for the management intervention, we did the following screening procedures. For TI-only circuits, screeners scheduled for a study visit the first one or two schools that passed the screening. For Management circuits, for the first school, we randomly selected whether or not to schedule the study visit. If the answer was “no” we then proceeded on to the next school on the list. Overall, five of the circuits were dropped because they did not have any schools that met the eligibility criteria for inclusion. Of the 140 remaining circuits, 70 each were assigned to either the enhanced management group or the non-enhanced management group. All CS from these two management groups were targeted for participation in the study;
4. Following the selection of schools, we randomly divided the non-management schools into Control and TI schools, stratified by district. Our final study sample, therefore, consists of two schools from the non-management circuits [one each for the control and targeted

¹ Since we conducted the baseline in May/June 2018, students in the study are in grades P5 and P6 during our study period [as at September 2018]. We did not include students in class 3 because not all upper primary schools have a class 3 class. In addition, the official language of instruction in class 3 is not English, and our partners elected to develop intervention resources in English language only.

² A shift system refers to a situation where two or more different schools use the same classrooms or school building based on a routine or schedule basis.

³ Multi-grade schools combine any of the primary 4 to 6 grades into a single classroom with a single teacher.

instruction] and one school from management circuits. This final sample consists of 210 schools made up of 140 schools and 70 schools from the non-management and management circuits, respectively;

Within each randomly sampled school:

- a. HTs from each sampled school (totaling 210) were automatically enlisted into the STARS project. This included 70 HT each from the control, targeted instruction, and targeted instruction-enhanced management.
- b. All P4-P6 teachers were purposively selected from each school (totaling 671 teachers). Specifically, we surveyed all teachers in the multi-stream schools.
- c. 15 students⁴ per class (30 per school) were randomly sampled for inclusion into the STARS project. If a particular class had more than 15 students, only 15 were randomly sampled. If less than 15 students, all students were automatically included. The total number of students enlisted into the STARS project was 5,904 students. At baseline, 5,904 students were recruited because of the unequal enrollment of children across primary 4 and primary 5. Students were randomly sampled using a SurveyCTO randomization program. If a school was multi-stream, we randomly selected one stream and only randomly selected students within one stream.

2.6 Outcome Measures

Our primary outcome measures are:

1. Did schools implement targeted instruction?
2. Did student learning improve, as measured by performance on math and English achievement tests?
3. Did teacher, HT, or pupil attendance improve?
4. Do CSs and HT management practices improve, as measured by:
 - a. Did the CS visit the study schools?
 - b. Did the CS or HT conduct effective classroom observations?
 - c. Did perceptions of effectiveness and mentorship by subordinates improve?

⁴ Only students who were present on the day of the visit and were listed on the classes 4 and 5 registers were sampled and included in the study.

Our secondary outcomes measures include:

1. Student aspirations and learning enjoyment: Do students enjoy math and English more?
2. Does teacher engagement with students in the classroom improve?
3. Does teacher burnout decrease and motivation improve?

2.7 Baseline Data Collection

2.7.1 Baseline questionnaires and content

Our baseline data collection involved surveys of four respondent types: 1) CS, 2) HT, (3) teachers, and (4) students. Notably, we used eight survey instruments for the baseline survey. These were the School Observation Instrument, CS Questionnaire, HT Questionnaire, Teacher Survey, Student Questionnaire, Student Assessment for Mathematics, and Student Assessment for English. The Principal Investigators from the University of Delaware and the University of Massachusetts Boston designed the questionnaires. The draft questionnaires went through a series of iterations before they were finalized. This involved reviewing, piloting the questionnaires and providing feedback to the PIs, through which the iterations were done.

Piloting was done to pinpoint problem areas in the instruments, reduce respondent burden, and estimate the time to administer the instruments as well as assess the interview/assessment flow and wording of the items. Questionnaire piloting was done on the 2nd and 8th of May 2018 at three private and public schools located at urban and rural localities - Roman Ridge, Dodowa and Volivo - in the Greater Accra Region of Ghana. The schools were Jack and Jill School, Dodowa D/A Basic and Volivo Primary schools. These schools were chosen to reflect similar characteristics of schools in the 20 UNICEF districts.

All instruments were administered in English and therefore there was no need to translate into the local language. After finalizing the instruments, the questionnaires were programmed in SurveyCTO by IPA's programmers. All programmed questionnaires were bench-tested to ensure that all consistency checks and skip patterns were working as expected. The following presents the contents of each of the Baseline Questionnaires.

- a. *School Observation Instrument*: The School Observation Instrument included measures such as infrastructure presence/absence, quality of infrastructure, the capacity of infrastructure, and student strength;
- b. *School Screener Form*: The School Screener Form was essentially a screening questionnaire to determine the eligibility of schools for inclusion in the STARS project. Screening questions included whether the school is a public school, whether the school

runs a shift system, has multi-grade (i.e., combined classes), and enrollment in primary 4 to primary 6;

- c. *CS Questionnaire*: The Baseline CS Questionnaire was administered to each CS. In Ghana, CS have a role in the middle management and serve as the link between schools and the district education office. On average, each CS is responsible for overseeing roughly 8 to 10 geographically proximate schools. The content of the CS Questionnaire included demographic characteristics, previous education/experience, supervision and mentorship of school personnel, management practices in schools, current management skills and abilities, activities carried out by them in schools, perceptions about teaching and learning in schools, and work-related stress and burnout. The CS Questionnaire was administered through a phone survey;
- d. *HT Questionnaire*: The HT Questionnaire was administered to the HTs in the sampled schools through an in-person interview. The HT Questionnaire sought information on the demographic characteristics, previous education/experience, supervision and mentorship of school personnel, management practices in schools, current management skills and abilities, activities carried out by them in their school, perceptions about teaching and learning in schools, and work-related stress and burnout;
- e. *Teacher Survey*: The Teacher Survey was administered to pupil teachers in classes 4, 5, and 6 through an in-person interview. The Teacher Survey had such modules as demographic characteristics, previous education/experience, activities conducted by them during the school day and outside, perceptions about teaching and learning in schools, mentorship from HT and CS, perceptions about teacher effort and absenteeism, and work-related stress and burnout;
- f. *Student Questionnaire*: The Student Questionnaire was administered to students in classes 4, 5, and 6. The questionnaire captured information on the students' background characteristics and aspirations, students' feedback on classroom teaching, and enumerator information about the quality of students' uniform (as a proxy for socioeconomic status);
- g. *Student Assessment for Mathematics and English*: The Student Assessments for Mathematics and English were administered directly to the selected students in classes 4, 5, and 6. These instruments were adapted from the TCAI assessment tool, developed by the Curriculum Research and Development Division of the Ministry of Education. Additional standardized tests developed by the Ghanaian National Education Assessment Unit [for grades 4 and 6] and from the EGRA and EGMA tools were included.

2.7.2 *Field staff recruitment and training*

Field staff was selected for training following a competitive recruitment process that brought together a combination of IPA-experienced and newly hired enumerators. Different categories of field staff were selected and trained. These were field supervisors, team leaders, auditors, and interviewers/child assessors. A total of 64 field staff qualified for the fieldwork and were subsequently selected.

Different training sessions relating to Pilot Survey Training, Supervisor Training, and General Training were conducted. Training approaches focused on classroom teachings and field practice. Classroom teaching practices comprised presentations, questions and answers, group discussions, and role-plays. Field practice was organized on the 23rd of May 2018 across 10 schools in the Greater Accra Region. Pilot survey training was held on the 30th of April 2018 in the IPA office. Ten people were trained on the baseline questionnaires. Only eight surveyors were retained for the actual piloting. A one-day supervisor training was held for the 20 candidates competing for the field supervisor and team leader positions before the general training sessions. This was to help them have a general overview of the entire baseline survey and enhance their participation and support to the training facilitators. This training took place on 16th May, a day before the general training. Finally, the general training was conducted for 70 candidates competing for the surveyor positions. The general training was held from 17th to 23rd May 2018 (i.e., 4 days of classroom training and one day of school visit). The survey field team comprised of 64 field staff members. The entire field team consisted of 12 survey field teams with each team composing of one team leader and four surveyors. Additionally, there were 2 field supervisors and 2 auditors.

2.7.3 *Baseline coverage*

Table 1 presents the coverage rates for the instruments administered to the CSs, HT, teachers, and students. We reported an average coverage rate of 98% for all the data collection questionnaires. Hundred percent of the CSs in the sampled circuits participated in the baseline survey. Six percent more of the teachers were surveyed due to the multi-stream nature of some of the schools, that is, some of the schools had more than one stream and hence teacher. We surveyed or assessed about 94% of the children. The remaining 6% were not assessed because of non-availability of children, refusal, and some students being children with special needs.

Table 1: Baseline Coverage Rate by Instrument

Instruments	Target	Completed	Refused	% Coverage
School Observation Instrument	210	210	0	100%
CS Survey	145	145	0	100%
HT Survey	210	209	1	99.5%
Teacher Survey	633	671	0	106%
Student Questionnaire	6300	5894	6	93.5%
Pupil Assessment [English]	6300	5894	6	93.5%
Pupil Assessment [Math]	6300	5893	7	93.5%

2.7.4 Data collection challenges

The research team encountered the following challenges.

- a. Limited or no school resources to support evaluation activities. Majority of the schools have limited or no school resources such as extra classrooms, chairs, and tables, which hinders the setting up of a conducive environment with the necessary resources for the interviews or pupil assessment activities;
- b. Limited enrollment and pupil absenteeism in some schools. Some schools had less than 15 children per class and hence the team could not meet the target of 30 children per school. In addition, the data collection activities were hindered by the pupil absenteeism;

- c. Poor network Connectivity. There was poor network connectivity due to the remote nature of the localities in which the sampled schools were located. This hindered communication among the team members and the synchronization of data through the SurveyCTO platform;
- d. Difficulty in assessing circuits or schools. Poor road networks make transportation to circuits or schools extremely difficult and expensive in some cases.

Section 3: Baseline Findings and Validation Tests

3.1 Description of Data Characteristics

In this subsection, we describe the means across the various measures and respondents. The next subsection addresses the similarities between the various groups by treatment status, i.e. baseline balance.

3.1.1 *Circuit Supervisors*⁵

We restricted to the sample of 140 circuits included in the intervention; all CS were trained in TI but half also received the enhanced management training. Almost all CSs were male (91%) and had been a HT prior to being a CS (69%). CSs were on average 45 years old, had worked as a CSs for about 4 years after spending 14 years as a teacher.

CSs meet on average about 12 times per year with other CSs in their district and visited each school 5 times per year. They hold just under 3 staff meetings per year and about 5 teacher training. When asked about the particular schools in our sample, they had visited them about 8 times in the last school year.

When asked to self-evaluate, most CSs strongly agreed that they were effective mentors to teachers (64%), effective mentors to HT (61%), and provided constructive feedback to their teachers (53%). Despite these responses, only 31% thought that they were much more effective than other CSs in similar circuits.

CSs report minimal absences in the last 5 days (an average of 0.2 days). When reflecting on terms 1 and 2, they were absent for official reasons about 6 days, absent for personal reasons about 2.5 days, and absent because they did not feel like working about one-third of a day.

A majority of the CSs feel every day that they are a positive influence (86%), are energetic about their job (85%), are satisfied with their job (74%). Similarly, 49% are never mentally exhausted,

⁵ See Appendix: Table 2 for more details on the data collected during Baseline, including these averages disaggregated by treatment group.

although 16% report feeling mentally exhausted a few times per week or more often. Many also report that they are never fatigued (29%), although 33 % report fatigue a few times per week or more often.

3.1.2 *Head Teachers*⁶

We interviewed 209 HTs. As with the CSs, HTs were overwhelmingly male (84%). They were on average 42 years old, had been a HT for almost 7 years, and had been a teacher for around 10 years.

The average school had 12 teachers, 178 enrolled boys, and 166 enrolled girls. HT reported that the CSs had visited their school 3 times in the first two terms of the year and that they had met with their CSs 5 times. These five meetings are similar in number to the reports from the CSs.

HT report that their teachers almost always use a lesson plan (90% of the time). Only about half use data for student promotion decisions (51%). They hold 5 staff meetings per year, 4 teacher training per year. Just under half (47%) reward their teachers in some way for good performance. About two-thirds weekly observe teaching (63%), give suggestions (65%), and discuss problems with their teachers (66%). Most HT strongly agrees that they are effective mentors (57%) and that they provide constructive feedback (51%). They also mostly agree that their CSs are an effective mentor (58%), a similar number to the percentage of CSs who report being an effective mentor (61%). HT are less sure of their abilities than the CSs with only 12% reporting that they are more effective than other HT in similar schools.

HTs report that on average 20% of teachers are late each day. For their own absence, HTs report an average of 0.5 absences for themselves in the past 5 days. HT reported 7 officially excused absences during terms 1 and 2, about one day more than those reported by CSs for themselves. HT were absent for personal reasons 3 days during terms 1 and 2 and only absent 0.14 days because they did not feel like coming to work.

Similar to the CS, most HT feel that every day they are a positive influence (78%), very energetic about their job (83%), are satisfied with their job (71%). A large percentage report that they are never mentally exhausted (49%), while 12 percent report mental exhaustion a few times per week or more often (not shown). Almost one-third (33%) report that they never feel fatigued, although 25% report fatigue a few times per week or more often. Therefore, HT are not generally self-reporting substantial work stress or burn out, although some do.

⁶ See Appendix: Table 3 for more details on the data collected during Baseline, including these averages disaggregated by treatment group.

3.1.3 Teachers⁷

We reached 671 teachers. As with the CSs and the HT, most are male (74%). They are on average 31 years old and have been teaching 6 years, 3 in the current school.

Teachers report spending 5 hours per week on lesson plans at home and 0.65 hours per week on lesson plans at school. Fewer than one-third of teachers (23%) think that they are much more effective than other teachers in similar schools. They report having attended 1.5 teacher trainings, 5 staff meetings with the HT, and 2 staff meetings with their CSs. About 20% of teachers had received a reward for good teaching. HT observed teachers on average 8 times in the last school year. Only 13% feel they can approach their CSs for help with a problem while almost all (93%) would approach their HT. About 77% report that their HT provides feedback on teaching.

Teachers were late to work on average 0.4 times over the last 5 days and were absent about half a day over the last 5 school days. During terms 1 and 2, teachers reported being absent for personal reasons about 3 times and the same number (3) of absences for personal reasons. They report only 0.1 missed days because they did not wish to work.

Many teachers strongly agree that they feel valued (60%), although a smaller percentage strongly agree that their HT is helpful (35%), and that their CSs are helpful (30%). Most strongly agree that they are very confident in the classroom (70%) and are very supported by someone at their school (52%).

Teachers report lower levels of motivation for their work compared to their supervisors. However, most still agree that every day they feel that they are a positive influence (69%), are very energetic about their job (80%) and are satisfied with their job (65%). Many report that they are never mentally exhausted (66%) or fatigued (43%). Similar to the CS and HT surveys, however, there is variation among all these measures with some teachers reporting lack of support and high overall stress levels.

3.1.4 Students⁸

Students in our study completed tests in Math and English. We used a partially adaptive test where students all took the same initial questions, then based on their answers were then given either an additional set of harder or easier questions. We use item response theory to solve for their latent score and put all students on the same scale⁹. We then standardize this score to mean

⁷ See Appendix: Table 4 for more details on the data collected during Baseline.

⁸ See Appendix: Table 5 for more details on the data collected during Baseline.

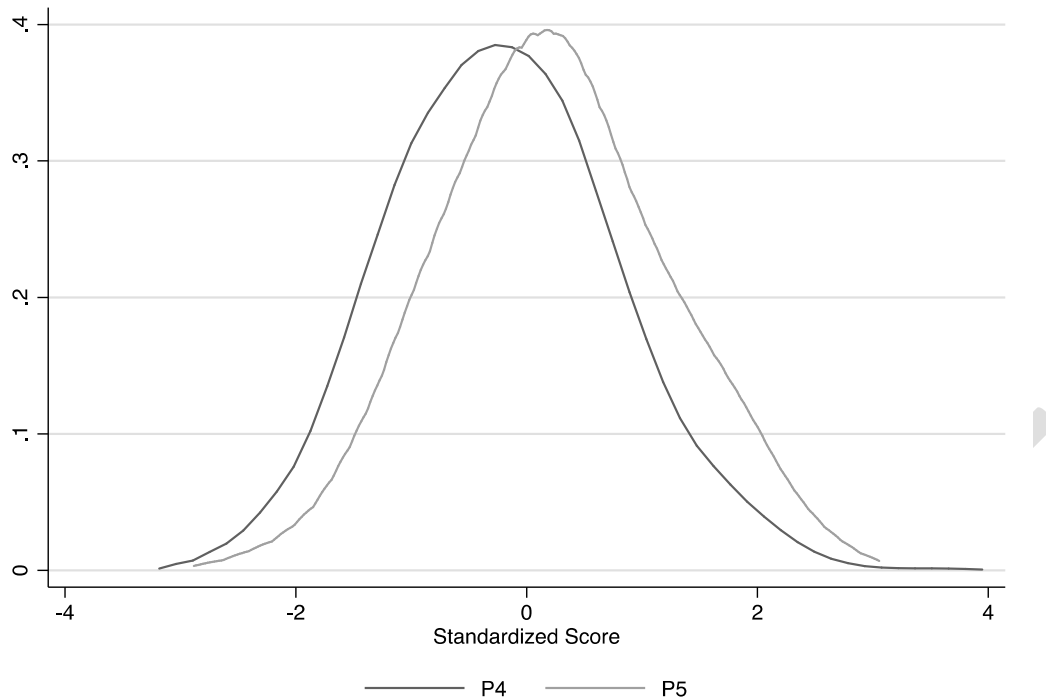
⁹ Item response theory presupposes a relationship between correct responses to a specific test question and a student's underlying latent ability. Correct responses to more difficult questions imply a higher latent ability than correct responses to easier questions. This is the same logic behind all adaptive tests, such as computerized standardized tests. For a more detailed description, please see Van der Linden and Hambleton (2013).

0 and standard deviation 1 within each subject. Therefore, by design, the average score of this test should be close to 0. We then combine these two subject scores to result in an aggregate latent ability score. Indeed, in Table 4, the average standardized test scores are approximately 0. Overall, approximately 22 percent of students scored well enough to take the more difficult tests in math or English.

Figure 2 graphs the distribution, by grade, of the combined latent ability measure. This figure highlights the substantial overlap in the distribution of student ability by grade. Although students in P5 have a higher mean, there are many P5 students who have a similar ability measure than P4 students. This motivates the reasoning behind targeted instruction: teaching students lessons that match their ability, instead of the official curriculum for their grade, should result in more effective teaching for all students.

For the demographic characteristics, just under half of our students are female (47%) and then are on average 12 years old. Students report about one-third of mothers are literate (37%) and just over half of their fathers are literate (54%). Almost all students want to attend senior high school (95%) and students report being absent about half a day in the previous week (0.6 days). 67% of students report liking their math class very much and 44% like English class very much.

Figure 2. Distribution of Combined Student Latent Ability by Grade



3.2 Baseline Equivalence on Outcome Measures

In addition to providing insight into means across the sample, we also tested for the equivalent of means across the treatment arms. Addressing the respondents in the same order as above, we start with CSs.

For CSs we collected 25 different measures, three of which are statistically different at the 10% level across the treatment arms, or 12%. This is higher than one would hope given our random assignment, but as we are randomizing across only 140 CSs and sometimes such imbalances occur. The four items that are different are that CSs in the management arm are more likely to assess themselves as an effective mentor to teachers (0.56 vs. 0.71) and HT (0.53 vs. 0.69), and have been absent in the last 5 days (0.09 vs. 0.27). That these differences do not always favor one group—those in the management arm are self-assess better mentors but also absent more often—is reassuring. Given these differences, we can control for them in the analysis of the follow-up data.

For the HTs we collected 31 different measures, three of which are statistically different between the three treatment types. HT assess their effectiveness as a manager differently across the three treatment arms with those in the management arm (67%) more likely than those in the control arm (57%) or those in the TI only arm (47%) to strongly agree that they are an effective manager. In the control arm, 73% of HT report that they discuss problems weekly, compared to 69% in the TI only arm and 55% in the management arm. The final difference is 41% of HT in the control arm

report their teachers to receive performance rewards, while 43% of HT in the TI only arm and 58 % of HT in the management arm report giving rewards.

For the teachers, we collected 30 different measures, three of which are statistically different between the arms at the 10% level.

For students, we tested students and asked basic demographic characteristics. Across these 11 measures, no measures are statistically different across the arms at the 10% level.

3.3 Risks to Impact Evaluation Design

Overall, the sample is balanced across the various treatment groups. Over the 97 different measures collected, 9 or just over 9% are statistically different at the 10% level, approximately what one would expect from a truly random sample.

One benefit of collecting baseline data is that we can control for any baseline differences in our estimation. Therefore, we do not see any threat to impact evaluation design based on data collected in the baseline.

As expected, one concern is spillovers of the management intervention onto non-treated CSs. On average CSs report significant interaction with one another within a district, although the variability of responses within a district indicates that there is potentially substantial measurement error. We will continue to monitor and include questions at endline regarding the sharing of materials across treatment groups. It should be stressed though, that any spillover will only potentially mitigate the impact of the management training and will still result in an unbiased estimate of the effect of TI on learning outcomes.

Given anecdotal evidence on a CSs, HT, and teacher stress and burn-out, we are somewhat surprised to find little evidence in the baseline data. Thankfully, relatively very few individuals demonstrate such issues. Therefore, the intervention having meaningful changes on the margins that we have measured is unlikely. Of course, some of these answers might be a hesitancy in telling a stranger anything negative. In the remaining follow-up rounds, we might explore alternative ways of measuring these outcomes.

3.4 Analysis and Lessons Learned from Baseline

Overall, the data that we collected appears to be of high quality. The self-reported absenteeism measures for the adult respondents are lower than collected through other studies' objective measures. We will be able to confirm the rates in our sample with two spot checks during the current (2018-2019) academic year.

HT appears to be already visiting classrooms but perhaps not providing feedback directly to the teachers. One margin in which we might see movement is the extent to which teachers feel supported at school, with only half of the teachers feeling that way now.

CONFIDENTIAL

Section 4: Intervention Implementation

4.1 Implementation of STARS Intervention

Preparation for the main roll-out of STARS started with revisions to all training and teaching materials with consideration of the key lessons from the pilot intervention. The table below summarizes the key activities that occurred prior to the roll-out of the STARS intervention.

Table 2. Timeline for Key Implementation Activities

Activity	Timeline
Regional and district orientations	18th - 19th July 2018
National TOT and master trainers' workshops	23rd July - 17th August 2018
Resource preparatory meetings for district training	20th - 24th August 2018
Teacher training on targeted instruction	27th - 31st August 2018
Enhanced management training for HT, and CSs	4th - 6th September 2018
School-based assessments, groupings, assessment verifications	11th - 21st September 2018
Targeted instruction implementation roll-out	24th September 2018

The Ghana Education Service led the orientation for the roll-out of STARS roll-out which took place in Kumasi from July 18 – 19th in 2018. The meeting saw about 95 percent of expected participants from regional and district level officers across all 20 UNICEF districts. The target audience for the orientation included regional and district directors, training officers and Assistant or Deputy Directors of supervision. Participants were introduced to the STARS intervention, the implementation approach, pilot lessons on targeted instruction and the school's environment, and the roles of regional, and district level officers. In these meetings, participants shared their opinions on key implementation challenges including inadequate teachers in schools, pupils' absenteeism, challenges with school-level data collection, recruiting of District Teacher Support Team (DTST) members for teacher training on Targeted Instruction. Directors acknowledged their key role in encouraging that the right trainers are recruited to train teachers, and trained teachers

stay at the current post for the next academic year. Overall, the orientation induced regional and district commitment towards the intervention.

The national core team and master training workshops were held in Kumasi and Koforidua to prepare adequate training capacity for the teacher training at the various districts. Overall, 24 national trainers were resourced to train 159 DTSTs. Their training followed a day resource preparatory workshop at their district to prepare for their District Teacher Training on Targeted Instruction.

The teacher training on targeted instruction came-off from 27th – 31st August. On average, eight (8) tests led by the District Training Officer (DTO) organized the targeted instruction training in each district. Training was successful in all 20 districts. Overall, about 80 percent of teachers in the research sample were trained. Reasons, why some of the HT, teachers, and CSs did not attend the training, included transfers from sampled schools to non-sampled schools, participation in the sandwich program, sickness, and travel. The two major challenge encountered were replacements of teachers who were on vacation degree programs popularly known as sandwich courses and insufficient materials for all training participants. Following the training, plans are being made to hold a mop-up training for all participants who missed the main teacher training as well as new teachers who have been newly posted to our implementation schools. This training is scheduled to come off from the 1st to 5th October 2018. Ghana Education Service, through its Basic Education Division, liaised with UNICEF to get materials printed and distributed to all districts/schools as a way of getting all relevant people to have access to the materials needed for implementation in schools.

The 2018 / 2019 academic year began on the 11th September 2018. School-based assessments and groupings have been conducted in the majority of the schools. The project has also set up an alternative system to run the Help Desk and the SMS component of the enhanced management design since the MSRC cannot be operational in the first term.

4.2 Intervention Piloting

The STARS intervention was piloted in two districts: Karaga and Asikuma-Odoben-Brakwa (AOB) from 18th June to 13th July 2018. The pilot tested both targeted instruction and aspects of the enhanced management approach of the STARS intervention. Prior to the pilot activities, the Ghana Education Service in collaboration with the key partners on STARS (UNICEF, NIB, NTC, NaCCA, and IPA) organized orientation meetings to orient regional and district officers about the STARS project and its objectives, and to solicit their input for the pilot roll-out in their districts.

Ten (10) schools participated in the pilot: eight (8) from AOB and two (2) from Karaga. All schools conducted assessments for grades 4, 5, and 6 using the STARS adapted ASER tool. Participating schools adjusted their timetables to make time for the targeted instruction lessons, 1 hour per

day, and 2 hours a week for English and 2 hours for Math. Most schools conducted targeted instruction lessons between 8:00 – 9:00 am. Schools mostly conducted English on Mondays and Tuesdays, and math on Wednesday's and Fridays. In cases where schools could not hold targeted instruction classes on the scheduled dates, targeted instruction classes were held on Fridays to make up for the lost schedules. On average, 80 percent of grade 4, 5, and 6 children participated in the pilot.

Activities of the pilot were monitored through regular field visits by the STARS technical team, officials from Pratham and JPAL within the four weeks of pilot implementation. Each monitoring team shared reports and key lessons were highlighted for improvement of materials, training, and implementation.

4.2.1 Observations from the pilot

From the pilot, it was observed that teachers and pupils are excited about the program. The intervention is increasing pupil willingness to come to school. Teachers are also recommending that the program is given more time to ensure lagging pupils catch up with the curricula for classroom learning.

- a. Teachers were excited and committed to the targeted instruction implementation wish they could implement targeted instruction approach in regular classrooms;
- b. Almost all targeted instruction lessons run over an hour. Aside from teachers needing more time to complete the plan, they did not have quick plans to revert to regular class activities after targeted instruction;
- c. Some irregular students who missed the school assessments and leveling were added to learning groups without being assessed;
- d. Group activities were not as engaging as expected. This was partly from the design of the teacher's guide and lack of enough teaching skills on this;
- e. Some schools are running multiple interventions and required careful planning run the interventions concurrently;
- f. Staff capacity for some schools is very low. This posted some managerial challenges for the school although this did not affect the intervention;
- g. Some school reported high teacher and pupil absenteeism as the main challenge for the intervention in their schools.

4.3 Key Lessons from Implementation

The following highlights key lessons from the implementation of the intervention.

- a. Targeted instruction lesson is affecting regular instructional hours. This may demand some clear guidelines on school time management;
- b. Schools undertaking other kinds of school interventions may require additional support from the CS;
- c. Teacher and pupil absenteeism can jeopardize the effectiveness of the intervention if not well managed. More stakeholder commitment is needed;
- d. School aggregate data are likely to record many errors if teachers and HT are not well trained (accurate record keeping is key);
- e. The intervention is likely to be affected in schools where time is poorly managed. Both schools did not have good time management skills;
- f. Student progression may be a challenge for teachers if clear guidance is not given on how teachers manage pupils who move within the term. Now, teachers strictly follow the schedule of the lesson plan;
- g. Teachers may do better at managing instructional hours if the TG is well referenced for all activities and teachers are trained in all activities during the targeted instruction training;
- h. HT and CS providing less support than expected. They will require more targeted instruction specific training on coaching, mentoring, and leadership.

Section 5: Conclusion and Next Steps

5.1 Conclusions

Our baseline was successful on many margins. We reached the targeted number of circuit supervisors, schools, and students. The data that we collected appears to be high quality and balanced across the various treatment arms. Through an analysis of the data, three lessons emerge.

First, our results suggest that student learning levels are indeed weakly correlated with grade level, with substantial overlap across the different grades. Thus, our baseline survey reinforces the need to better tailor pedagogy to learning levels instead of official curriculum based on grade level.

Second, our results suggest that based on self-reports by CSs, HT, and teachers, monitoring and interacting with management is occurring regularly. Teachers and HT generally report adequate levels of classroom observations and feedback receipt. While the levels might be sufficient, lack of student learning indicates that these status quo interactions are not adequate. This project will provide multiple independent observations of actual classroom practices to test for improved classroom practices.

Third, reported satisfaction with the status quo is high on some measures that we thought the intervention might improve, e.g. teachers already report feeling very energetic about their jobs. Therefore, we are currently reviewing these questions and plan to modify them for the endline.

Given ongoing challenges with poor student outcomes, our results suggest the need for all actors to be more efficient and effective in producing student learning gains. This intervention will test one possible way to improve the production of education in the classroom and through improved managerial support. As a result, STARS will provide lessons to both Ghana and other countries thinking about taking targeted instruction to scale.

5.2 Next Steps

The first Observational Study will begin in October 2018, with a second Observational Study planned for February 2019. The final rounds of data collection – Follow-up – will be conducted in May 2019. These data collection activities will involve interviewing the key stakeholders – HT, teachers, CSs, and students. Next implementation activities include organizing a mop-up training for teachers who could not participate in the targeted instruction training during vacation.

Appendix

Table 2: CS characteristics, by Treatment Group

	TI Schools (N=70)	TI +Management (N=70)	P-Value From F-Test of Equal Means
	(1)	(2)	(3)
<u>Panel A: Demographic Characteristics</u>			
Male	0.91	0.90	0.77
Age	46.26	44.64	0.19
Years CS	4.41	4.07	0.51
Years As Teacher	14.39	13.90	0.71
Ever HT	0.69	0.69	1.00
Years As HT	6.60	5.83	0.47
<u>Panel B: Management and Work Practices</u>			
Num. Meetings with Other CS	12.41	12.95	0.80
Num. Meetings with Each HT	4.70	5.84	0.14
Num. Staff Meetings	2.34	2.90	0.26
Num. Teacher Trainings	4.93	5.38	0.59
Number of Visits to School 1	7.78	7.80	0.97
Number of Visits to School 2	8.04	.	--
<i>Self-Evaluative Measures</i>			
I am an effective mentor to teachers (strongly agree)	0.56	0.71	0.05
I am an effective mentor to head teachers (strongly agree)	0.53	0.69	0.05
I am a much more effective CS than others	0.37	0.26	0.15
I provide constructive feedback (strongly agree)	0.50	0.56	0.50
<u>Panel C: Absence Measures</u>			
Num. absences, last 5 days	0.09	0.27	0.04
Num. officially excused absences during terms 1 and 2	5.76	7.22	0.13
Num. absences for personal reason during terms 1 and 2	2.63	2.49	0.88
Num. absences because not wish to work during terms 1 and 2	0.20	0.55	0.16
<u>Panel D: Motivation Measures (Higher is more motivated)</u>			
Positive Influence (every day)	0.87	0.85	0.615
Very Energetic (every day)	0.85	0.86	0.834
Job Satisfaction (every day)	0.76	0.71	0.493
Never Mentally Exhausted	0.49	0.52	0.736
Never Fatigued	0.26	0.32	0.489

Notes: P-Values in Column 3 use robust standard errors

Table 3: HT characteristics, by Treatment Group

	Control Schools (N=70) (1)	TI Only (N=70) (2)	TI +Management (N=69) (3)	P-Value From F-Test of Equal Means (4)
<u>Panel A: Demographic Characteristics</u>				
Male	0.81	0.86	0.86	0.75
Age	42.90	40.67	42.84	0.21
Years as HT	6.81	6.03	7.64	0.22
Years as Teacher	10.06	8.67	9.99	0.34
<u>Panel B: School and CS Characteristics</u>				
Num. Teachers	11.39	12.07	12.23	0.44
Num. Boys Enrolled (Total School)	170.16	184.09	179.68	0.57
Num. Girls Enrolled (Total School)	161.00	177.93	158.50	0.48
Num. CS Visits	3.03	3.46	3.20	0.62
Num. Times Meet with CS	5.07	4.96	5.30	0.90
<u>Panel C: Management and Work Practices</u>				
Teachers Always Use a Lesson Plan	0.87	0.90	0.93	0.54
Use Data for Student Promotion	0.47	0.51	0.55	0.65
Number of Staff Meetings	5.30	5.13	5.62	0.38
Number of Teacher Trainings	4.49	4.19	4.18	0.81
Teachers receive rewards for performance	0.41	0.43	0.58	0.10
Observe teaching weekly	0.54	0.69	0.67	0.18
Give suggestions weekly	0.61	0.63	0.72	0.31
Discuss problems weekly	0.73	0.69	0.55	0.08
<i>Self-Evaluative Measures</i>				
I am an effective mentor (strongly agree)	0.57	0.47	0.67	0.06
My CS is an effective mentor (strongly agree)	0.54	0.63	0.57	0.57
I am a much more effective HT than others	0.10	0.13	0.13	0.81
I provide constructive feedback (strongly agree)	0.51	0.51	0.52	0.99
<u>Panel D: Absence Measures</u>				
Pct. Of Teachers Late	0.22	0.21	0.18	0.46
Num. absences, last 5 days	0.46	0.54	0.52	0.86
Num. officially excused absences during terms 1 and 2	7.96	6.26	8.03	0.14
Num. absences for personal reason during terms 1 and 2	4.51	2.21	2.87	0.16
Num. absences because not wish to work during terms 1 and 2	0.10	0.26	0.07	0.47
<u>Panel E: Motivation Measures (Higher is more motivated)</u>				
Positive Influence (every day)	0.79	0.73	0.83	0.39
Very Energetic (every day)	0.8	0.86	0.81	0.63
Job Satisfaction (every day)	0.79	0.7	0.65	0.20
Never Mentally Exhausted	0.56	0.44	0.48	0.39
Never Fatigued	0.34	0.3	0.33	0.85

Notes: P-Values in Column 4 use robust standard errors.

Table 4: Teacher characteristics, by Treatment Group

	Control Schools (N=217) (1)	TI Only (N=226) (2)	TI +Managemen t (N=228) (3)	P-Value From F- Test of Equal Means (4)
Table 4: Teacher characteristics, by Treatment Group				
<u>Panel A: Demographic Characteristics</u>				
Male	0.74	0.75	0.73	0.91
Age	31.64	31.57	30.96	0.58
Years Teacher	6.17	6.13	5.95	0.94
Years Teacher, This school	3.09	3.47	3.29	0.48
<u>Panel B: Teaching Practices and Views</u>				
Hours Lesson Plan at Home	4.99	5.65	4.80	0.22
Hours Lesson Plan at School	0.76	0.64	0.56	0.54
I am a much more effective teacher than others	0.23	0.22	0.24	0.91
Number of Teacher Trainings	1.42	1.59	1.51	0.80
Number of Staff Meetings (by HT)	5.11	5.08	5.35	0.79
Number of Staff Meetings (by CS)	2.00	1.82	1.87	0.55
Received Performance Reward	0.20	0.28	0.22	0.18
Number of times HT observed Classroom	9.08	8.83	7.93	0.78
Approach CS for problems	0.12	0.11	0.17	0.21
Approach HT for problems	0.94	0.91	0.93	0.61
HT gives feedback on teaching	0.74	0.77	0.79	0.58
<u>Panel C: Absence Measures</u>				
Times Late to work, last 5 days	0.44	0.31	0.42	0.13
Num. absences, last 5 days	0.46	0.51	0.47	0.91
Num. officially excused absences during terms 1 and 2	2.48	2.50	2.74	0.83
Num. absences for personal reason during terms 1 and 2	2.47	3.36	3.26	0.13
Num. absences because not wish to work during terms 1	0.06	0.16	0.10	0.19
<u>Panel D: Support Measures (Higher is more supported)</u>				
Feel Valued (Strongly Agree)	0.62	0.66	0.54	0.02
HT Helps (Strongly Agree)	0.30	0.34	0.42	0.06
CS Helps (Strongly Agree)	0.26	0.33	0.30	0.30
Very Confident in Classroom	0.69	0.71	0.70	0.90
Very Supported, Someone at School	0.52	0.57	0.49	0.32
<u>Panel D: Motivation Measures (Higher is more motivated)</u>				
Positive Influence (every day)	0.62	0.73	0.71	0.07
Very Energetic (every day)	0.79	0.81	0.8	0.87
Job Satisfaction (every day)	0.62	0.68	0.65	0.54
Never Mentally Exhausted	0.66	0.7	0.69	0.64
Never Fatigued	0.43	0.38	0.38	0.48

Notes: P-Values in Column 4 cluster standard errors at the school level.

Table 5: Child Characteristics, by Treatment Group

Table 5: Child characteristics, by Treatment Group

	Control Schools (N=2032) (1)	TI Only (N=1932) (2)	TI +Managemen t (N=1929) (3)	P-Value From F- Test of Equal Means (4)
<u>Panel A: Standardized Scores</u>				
Math	0.01	-0.02	0.01	0.91
English	-0.01	0.00	0.01	0.99
Total	0.00	-0.01	0.01	0.99
<u>Panel B: Demographic Characteristics</u>				
Female	0.47	0.46	0.48	0.78
Age	12.17	12.05	12.12	0.74
Mother Can Read/Write	0.37	0.40	0.35	0.42
Father Can Read/Write	0.54	0.55	0.53	0.93
Want to attend Senior High School (SHS)	0.95	0.95	0.95	0.86
Days Absent, Last 5 Days	0.57	0.58	0.56	0.92
Likes English class very much	0.44	0.43	0.41	0.64
Likes Math class very much	0.67	0.64	0.62	0.11

Notes: P-Values in Column 4 cluster standard errors at the school level.

CONFIDENTIAL

References

- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics* 122 (3), 1235–1264.
- Bloom, N., Schankerman, M., & van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81, 1347-1393.
- Mano, Y., Iddrisu, A., Yoshino, Y., & Sonobe, T. (2012). How can micro and small enterprises in Sub-Saharan Africa become more productive? The impacts of experimental basic management training. *World Development*, 40 (3), 458-68.
- Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2018). How to improve teaching practice? Experimental comparison of centralized training and in-classroom coaching. Working Paper RISE-WP-18/024. Retrieved from https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-024_Cilliers_TeachingPractice.pdf
- Duflo, E., Dupas, P., & Kremer, M. R. (2011). Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. *American Economic Review* 101 (5), 1739–1774.
- Ministry of Education. (2016). Ghana 2016 National Education Assessment report of findings. Retrieved from https://globalreadingnetwork.net/sites/default/files/eddata/2016%20NEA%20Findings%20Report_17Nov2016_Public%20FINAL.PDF
- Pritchett, L. (2013). *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development; Baltimore: Brookings Institution Press.
- United National International Children's Emergency Fund. (2012). All children in school by 2015: Global Initiative on Out-of-School Children. Ghana Country Study. Retrieved from <http://unesdoc.unesco.org/images/0022/002213/221302e.pdf>
- UWEZO. (2014). Are our children learning? Literacy and numeracy in Kenya 2014. Twaweza East Africa, Nairobi.
- van der Linden, Wim J, and Ronald K Hambleton. (2013). *Handbook of modern item response theory*. Springer Science & Business Media.