STRENGTHENING TEACHER ACCOUNTABILITY TO REACH ALL STUDENTS (STARS)

Milestone 4: Impact Analysis Report

Prepared for the World Bank Group, the Strategic Impact Evaluation Fund - Contract No.: 7180606

Sabrin Beg, Anne Fitzpatrick, Adrienne Lucas, Edward Tsinigo, and Henry Atimone[1]

December 2019 (revised January 2020)

IPA Ghana

[1] Beg: University of Delaware. Fitzpatrick: University of Massachusetts, Boston. Lucas: University of Delaware, NBER, J-PAL, and CGD. Tsinigo and Atimone: Innovations for Poverty Action, Ghana.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BED	Basic Education Division
CAPI	Computer Assisted Personal Interviewing
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
GES	Ghana Education Services
IPA	Innovation for Poverty Action
J-PAL	Abdul Latif Jameel Poverty Action Lab
MOE	Ministry of Education
NaCCA	National Council for Curriculum and Assessment
NIB	National Inspectorate Board
NTC	National Teaching Council
P4	Primary 4
P5	Primary 5
P6	Primary 6
STARS	Strengthening Teaching Accountability to Reach All Students
T2	Treatment 2 (Targeted Instruction Only)
T3	Treatment 3 (Targeted Instruction and Enhanced Management)
TCAI	Teacher Community Assistant Initiative
TI	Treatment 1 (Control Group)
UNICEF	United Nations Children's Fund

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 project. Low adherence to this intervention and other interventions that have been proven effective 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 properly implementing the project. The Strengthening Teaching Accountability to Reach All Students (STARS) project was designed to train teachers on targeted instruction and improve the support that head teachers and circuit supervisors provide for those teachers. Through a randomized controlled trial (RCT), this project (a) tested the effect of training teachers on targeted instruction and (b) tested whether additional management training of head teachers (i.e. school principals) and circuit supervisors (i.e. middle-level management responsible for a subset of schools within a district) increased the quality of implementation of targeted instruction and student outcomes. This study worked within the Ghana education system to improve educational outcomes. Ghana Education Services (GES), National Council for Curriculum and Assessment (NaCCA), and the National Inspectorate Board (NIB) designed the materials and trained the teachers, head teachers, and circuit supervisors.

In May and June 2018, prior to the end of the 2017-2018 academic year, Innovation for Poverty Action (IPA) enumeration teams visited 210 schools, interviewing 209 head teachers and 671 primary (P) 4 through 6 teachers and interviewing and testing 5,893 P4 and P5 pupils in both English and math. The Milestone 2: Baseline Report summarizes those findings. In August 2018, before the start of the 2018-2019 academic year, existing GES personnel trained teachers, head teachers, and circuit supervisors. During the 2018-2019 school year, IPA enumeration teams conducted two unannounced observational studies, i.e. a spot check, and one full achievement follow-up.

We find that the two interventions were statistically indistinguishable from each other in most cases. They increased student learning on a combined English and Math test by about 0.11 standard deviations. These increases were likely not related to increased student attendance as students in the TI-only arm had lower attendance at the spot checks by about 3 percentage points. Teachers in the treatment arms were no more likely to be present on school grounds but were about 11 percentage points more likely to be in the classroom. Teachers appeared to implement the program with greater fidelity than in the previous TCAI study with over 25 percent of them using STARS teaching and learning materials. Further, they divided students by learning level instead of grade level in about 60 percent of spot checks. Head teachers were about 12 percentage points more likely to be present and according to teachers they performed about 0.7 more classroom

observations of at least 30 minutes and were about 12 percentage points more likely to provide feedback to teachers.

1. INTRODUCTION

1.1 CONTEXT

Many Sub-Saharan African countries have made considerable progress towards the achievement of Goal 4 of the United Nations Sustainable Development Goals that encourages countries to "Ensure inclusive and quality education for all." Evidence of this progress includes the increasing investment in promoting access to primary education. As in many countries in sub-Saharan Africa, primary enrolment rates in Ghana have increased substantially in recent years, but students' learning levels have not matched this progress. For example, the United Nations Children's Fund (UNICEF) Global Initiative 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 school 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 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 outcomes in developing countries is targeted instruction - teaching at the knowledge level of the student instead of grade level.

Previous research in Ghana and India demonstrated that targeted instruction is a cost-effective strategy to increase student learning (Banerjee et al., 2007; Banerjee et al., 2017; Duflo et al., 2019). The study in Ghana, known as the TCAI is particularly relevant to the present study. TCAI revealed that targeted instruction improved learning outcomes, although only a fraction of teachers implemented the project. These results suggest that with weak accountability, monitoring, and supervisory support for teacher-led projects, 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 project as designed. Thus, additional research and reflection could 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 5% of teachers correctly implemented it during unannounced spot-checks (Duflo et al., 2019). Strong accountability, monitoring, and supervisory support could enhance the broad take-up of the targeted instruction project. 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) were both more effective at improving student test scores than traditional pre-term teacher training projects. 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 Ghana Ministry of Education (MoE) and the GES. UNICEF has been supporting the MoE and 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.

This impact analysis report is based on the evaluation and implementation activities of the STARS project. The report provides evidence on the success of the targeted instruction and possible scale-up as well as developing evidence-based policies toward improving the quality of basic school education in Ghana and beyond.

1.2 THE STARS PROJECT

Focus group discussions with teachers, head teachers, and circuit supervisors following TCAI identified that weak management policies were a contributing factor to the low take-up of targeted instruction. Therefore, STARS was developed to replicate and improve upon the success of TCAI by including management support to a subset of schools. The STARS project encouraged fidelity of implementation of targeted instruction by leveraging the role of existing education personnel, i.e. teacher managers. The project built on the capacity of head teachers and circuit supervisors by improving their coaching and mentoring skills. Hence, these

supervisors supported teachers with monitoring resources to implement the targeted instruction pedagogy with fidelity. The project explored (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.

2. INTERVENTION AND IMPLEMENTATION

2.1 INTERVENTION

The STARS project comprised of two key interventions, namely, targeted instruction and enhanced management.

2.1.1 TARGETED INSTRUCTION

Targeted instruction under the STARS project focused on equipping the classroom teachers from primary (P) 4 to 6 with the requisite skills to teach at the knowledge level of the child. The focus was to help pupils in the targeted grades who were performing below grade level in English and Math acquire the fundamental skills to perform at grade level. The project built on the capacity of teachers, head teachers, and circuit supervisors on how to deliver the targeted instruction pedagogy. At the beginning of the school year, all pupils in P4 to P6 were tested in Mathematics and English using a basic ASER[2]-type tool and grouped into 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 delivered targeted instruction to pupils based on the learning levels. Teachers used formative assessments to track the progress of pupils. Students were 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 provided an opportunity for the teacher to track students' progress and move students who make progress into higher levels. During the targeted instruction hour, teachers engaged children in whole class, group, and individual activities.

1. Class activities: The whole class activities were designed to ensure that students were engaged in different activities, come out of their inhibition and feel confident to converse freely in peer groups and with the teacher.

[2] The Annual Status of Education Report

2. Group activities: The group activities were designed to facilitate student engagement with their assigned activities.
3. Individual activities*:* These were designed to promote independent work of students regarding reading and writing through their participation in certain activities including creating 'mind-maps' and 'reading a simple paragraph'.

2.1.2 ENHANCED MANAGEMENT TRAINING

The circuit supervisors and head teachers received additional training on the best practices to mentor and support teachers as they implemented the project. Circuit supervisors and head teachers received a succinct resource manual on how to perform specific coaching and mentoring support activities to teachers. In addition, circuit supervisors and head teachers received positive reinforcement SMS text messages directly from GES that included tips on how to be a better mentor/coach and perform effective monitoring and classroom observations. Also, teachers, circuit supervisors, and head teachers had access to a GES-run Help Desk where they could call or text to receive prompt feedback on any issues that they are facing.

2.2 COLLABORATION AND COORDINATION

The STARS project was a partnership among many institutions - MoE (through the following agencies: NIB, National Teaching Council (NTC), and NaCCA), GES, UNICEF, and IPA.

Under this partnership, MoE and GES leveraged its existing structures to implement the project in selected schools. GES as the implementing organization on the STARS project led the implementation of the project by setting up structures and systems for successful implementation. The Basic Education Division (BED) of GES supported NaCCA, NIB, and NTC to develop relevant materials needed for the successful implementation of the STARS project. Specifically, BED supported the development of the TI materials and circuit supervisor and head teacher Quick Reference Guides. BED participated in all partner engagement meetings including the quarterly core group meetings.

The National Council for Curriculum and Assessment led the development of TI materials. NaCCA constituted a Resource Development Team 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 for P4 to P6.

The National Inspectorate Board was responsible for the development of the circuit supervisors' and head teachers' intervention materials (manuals, content for text message reminders, etc.). NIB managed logistics and oversaw the manual development process, provided status updates

to stakeholders, and submitted final copies of manuals to the STARS Core Team ahead of project implementation.

The National Teaching Council was responsible for the development of training materials and training of teachers on the targeted instruction pedagogy. As part of its role, NTC constituted a team of core trainers who were responsible for (a) developing training guides and facilitator manuals and (b) conducting a train the trainer session for district training teams. NTC supervised the base intervention training in all districts and managed all the logistics associated with carrying out these tasks.

UNICEF provided technical and financial support, whereas IPA provided technical support in the intervention design and supported the principal investigators to conduct the evaluation.

2.3 KEY INTERVENTION MILESTONES AND TIMELINESS

Preparatory works leading to the implementation of the STARS projects started during the 2017-2018 academic year. The following highlights the key milestones in the preparatory work leading to the implementation activities.

In collaboration with UNICEF, NIB, NTC, NaCCA, and IPA, GES organized orientation meetings for regional and district directors, training officers and Assistant or Deputy Directors of supervision about the STARS project. The orientation meetings, which took place in Kumasi from 18th to 19th July 2018 had about a 95% attendance rate from the regional and district level officers across all 20 UNICEF districts. The orientation meeting focused on the overview of the STARS project, intervention, implementation approach, pilot lessons on targeted instruction and the school's environment, as well as the roles of regional, and district level officers.

Prior to the STARS rollout, the interventions were piloted in two districts: Karaga and Asikuma-Odoben-Brakwa from 18th June to 13th July 2018. The intervention piloting exercise saw the participation of 10 schools: 8 from Asikuma-Odoben-Brakwa and 2 from Karaga. All schools conducted assessments for grades 4, 5, and 6 using the STARS adapted ASER tool. On average, 80 percent of grade 4, 5, and 6 students in each school participated in the pilot. A joint STARS technical team monitored the piloting activities. This included officials from Pratham[3] and J-PAL. Feedback from the piloting informed the revisions to all training and teaching materials.

A one-day resource preparatory workshop was organized in each participating district to prepare for their district teacher training on targeted instruction. The national core team and

[3] Pratham is an innovative learning organization created to improve the quality of education in India.

master training workshops were then held for 24 national trainers in Kumasi and Koforidua to build their capacity for the teacher training at the various districts.

2.4 IMPLEMENTATION STATUS

The STARS interventions were implemented within one academic year, i.e., the 2018-2019 academic year. The implementation activities were conducted throughout terms 1 to 3 of the 2018-2019 academic year. These activities were monitored by a combined team of national and district education officials as well as IPA and UNICEF.

2.4.1 TARGETED INSTRUCTION

The main training of P4-P6 teachers, head teachers, and circuit supervisors on targeted instruction occurred from 27th to 31st August 2018. A mop-up training was conducted from 1st to 5th October 2018 for participants who missed the original training and the new teachers who were posted to the treatment schools. About 80% of the targeted 700 teachers and head teachers participated in the training for the targeted instruction. About 99% and 88% of the teachers and head teachers, respectively, participated fully in the 5-day training program. Non-attendance was largely due to (a) transfers of teachers from sampled schools to non-sampled schools, (b) teachers' involvement in the sandwich courses [4](c) sickness, and (d) travel. Schools received targeted instruction teaching and learning materials. However, term 1 training was confronted with the challenge of insufficient materials for trainees. In fact, about 60% of the 78 headteachers (i.e., 87 schools) who reported challenges associated with the implementation of the targeted instruction indicated insufficient materials for trainees. The inadequacy of these materials varied across schools and districts. For example, some schools reported limited supply of Teachers' Guide while others lacked readers, manila cards, word cards, sentence cards, number wheel, markers and or masking tapes and pair of scissors to develop some TLMs. Consistent with their training, we observed that some teachers used local materials to prepare teaching and learning materials for the implementation of the targeted instruction.

Refresher trainings were conducted during terms 2 and 3 of the 2018-2019 academic year to remind the teachers and head teachers about the targeted instruction pedagogy and to address any challenges they were facing with implementation. Unlike in term 1, adequate measures in terms of organization and communication were put in place to ensure maximum participation

[4] This is a training course with alternate periods of formal instruction and practical experience.

by all targeted teachers and headteachers. Due to these measures, the number of teachers and headteachers that participated in the refresher trainings increased significantly compared to the main training. During terms 2 and 3, the insufficiency of materials was considerably minimized.

Following the training for the implementation of the targeted instruction, the main school-based assessments, groupings, and assessment verifications were conducted from the 11th to 21st September 2018 for term 1. Mid-term assessments were conducted during terms 2 and 3. The assessment and grouping were done by the teachers. The head teachers were responsible for auditing the assessment and grouping process. The circuit supervisors did not play any active role in the assessment and grouping. The assessments led to the grouping of pupils into three learning levels.

2.4.2 ENHANCED MANAGEMENT

Training of head teachers and circuit supervisors for the enhanced management intervention was conducted from 4th to 6th September 2018 for term 1. The enhanced management training equipped the head teachers and circuit supervisors with efficient managerial skills - how to mentor, support, and observe teachers, and collaborate with them as a cohesive unit to be pedagogical leaders and improve student outcomes. The training also empowered 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.

No training evaluation was conducted for terms 1 and 2. During the term 3 training, however, participants were evaluated to allow them to independently assess the training content and the knowledge gained. The participants were quizzed on their understanding of key components of the implementation of the targeted instruction. More than 50% of the participants had 4 out of 5 questions correctly. Over 85% had at least 3 out of the 5 questions correct.

3. EVALUATION METHODS

3.1 EVALUATION OBJECTIVES

The STARS project leverages existing personnel, systems, and structures in the MoE and GES to improve student achievement and answer the following questions:

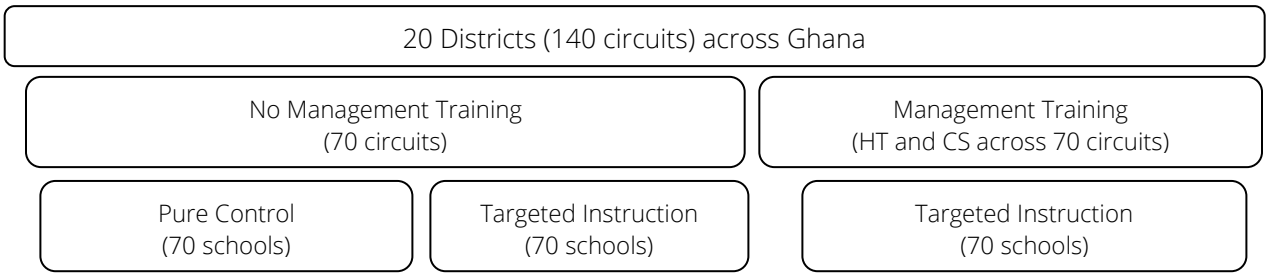
1. Does targeted instruction improve student test scores in upper primary grades?

2. Can monitoring, managing, and coaching performed by existing circuit supervisors and head teachers increase the likelihood that teachers implement targeted instruction?
3. Do enhanced monitoring, managing, and coaching lead to higher learning gains?

3.2 RANDOMIZATION DESIGN

This randomized controlled trial was conducted in the 20 districts in Ghana that UNICEF supports. This study operated in 140 circuits (groups of schools) within these districts.[5] 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). Notably, we first randomized each circuit to either receive the management training or not. Then within each management circuit, we randomly selected one school to receive targeted instruction and enhanced management training. Within each non-management circuit, we randomly selected two schools - one received targeted instruction training and the other did not. The study design is pictured in Figure 1.

Figure 1. The Study Design



This scheme results in three treatment arms.

1. Treatment 1 [TI - Control Group]. P4-P6 teachers and head teachers continue as usual receiving no additional training or resources. Due to project design, circuit supervisors in this arm received targeted instruction training.[6] Since the previous iteration of targeted instruction in Ghana suffered from a lack of and not overzealous implementation (Duflo, Kiessel, and Lucas 2019), circuit supervisors were unlikely to encourage schools in Treatment 1 to implement targeted instruction without materials or training, but it is possible. Treatment 1 comprised of 70 schools.

[5] These districts contain 145 total circuits. Five circuits were excluded. See details below.

[6] As is the norm in Ghana, CSs are included in the targeted instruction teacher trainings. Since one school in each circuit receives the targeted instruction treatment (Treatment 2), all CSs would be exposed to the training.

2. Treatment 2 [T2 – Targeted Instruction Only]. P4-P6 teachers, head teachers, and circuit supervisors all received targeted instruction training. Schools received targeted instruction teaching and learning materials. Head teachers and circuit supervisors did not receive enhanced management training. T2 consisted of 70 schools.
3. Treatment 3 [T3 - Targeted Instruction and Enhanced Management]. P4-P6 teachers, head teachers, and circuit supervisors received targeted instruction training. Schools received targeted instruction teaching and learning materials. Head teachers and circuit supervisors also received enhanced management training. T2 consisted of 70 schools.

This study design was similar to a fully cross-randomized design, although we did not have any schools that received just the management training intervention (no targeted instruction training). Our design has the advantage of allowing us to measure the impact of targeted instruction alone compared to business-as-usual, and the impact of T3 compared to business-as-usual. It also allowed us to compare the differential effect of the management intervention with T2. There were no control schools in the treatment circuits because we did not anticipate that the management training intervention, which is an enhancement of the base targeted instruction training, would substantially affect outcomes in the absence of the base targeted instruction intervention. In all control schools, the circuit supervisors received training as the sample included nearly the universe of circuits (and circuit supervisors) in the 20 districts. Dividing the circuits into 3 groups would have lowered the number of circuits in each treatment group and limited the power of our study.

3.3 EVALUATION APPROACH AND METHODS

3.3.1 SAMPLE SELECTION

Participants were sampled using a three-stage sampling design. The STARS study purposively selected the 20 UNICEF supported districts. As UNICEF funded part of the implementation of this project, the project only operated in the 20 districts in which UNICEF already had an established presence. We combined two different administrative data sources to create the universe of schools within the 20 districts.

Using the administrative data, we established a set of eligibility criteria for including the circuits and study participants. Only schools that met the following criteria across grades 4 through 6 were eligible for the STARS study: (a) non-zero enrollment, (b) non-shift, and (c) non-multi-grade. Relatedly, circuits that had no schools that met the eligibility criteria were excluded.

The following outlines the detailed procedures used to select the circuits, schools, and research participants for this study (see Appendix C). Our sampling strategy ensured that the circuit and school selection were identical in the two groups.

1. The 20 UNICEF-supported districts contained 147 circuits. Two districts – one each from northern and southern Ghana - were selected for piloting the intervention implementation, leaving 145 eligible circuits for inclusion in the study. The 145 circuits were randomly divided into two groups stratifying at the district level: the non-enhanced management arm (Group 1) and the enhanced management arm (Group 2).
2. Randomly ordered lists of all schools within each circuit were created using EMIS data on the 20 districts. Within each circuit, we phoned the head teachers of the schools in this random order and administered a screening questionnaire to determine their eligibility.
    a. The first two eligible schools from the Group 1 circuits were selected [one after the other] for the study. Once these two eligible schools were determined, the phone screening process stopped. These two schools were then randomly assigned to either the T1 or T2 experimental group. If a circuit had just one eligible school, it was randomly assigned to the TI or T2 experimental group. Circuits with no eligible school were excluded from the study.
    b. The first eligible school from the Group 2 circuits was selected for the study with a 50% probability (using a random assignment). The team developed a SurveyCTO randomization code [based on a 50%-50% chance] to determine whether a school should be selected. The first school within the Group 2 circuits that met the eligibility criteria was selected for the study using the SurveyCTO code.
    c. These strategies resulted in the exclusion of 3 of the 143 circuits from the study, as they did not have any schools that met the inclusion criteria. Of the 140 remaining circuits, 70 each were in Group I and Group 2. The experimental sample of 210 schools included two schools from the Group 1 circuits (140 total) and one school from the Group 2 circuits (70 total).
3. The selection of the research participants – circuit supervisors, head teachers, teachers, and students - was contingent on the inclusion of their respective schools or circuits in the study. The study population at baseline comprised of (a) 140 circuit supervisors of the 140 study circuits, (b) 209 head teachers of 209 out the 210 study schools[7], (c) 671

[7] One head teacher allowed the study in her school but refused to be interviewed because she had been interviewed recently for another IPA study - Quality Preschool for Ghana.

teachers[8] from primary 4 to primary 6 in the 210 study schools, and (d) 5893 students[9] in primary 4 to primary 5 at the end of the 2017-2018 school year.

3.4 RESEARCH INSTRUMENTS

Different research instruments were used to collect relevant information on the research participants across the various data collection waves. The development of these instruments followed a rigorous process of reviewing the literature on existing instruments for measuring indicators relating to the research questions. The review process led to the research team using a combination of the following in designing the research instruments: adopting existing tools, adapting existing tools and designing project-specific tools to measure key indicators that were not captured in previous instruments. These research instruments were thoroughly piloted to ensure that they were aligned to the context and adequately measured the indicators we sought to measure. The observations from the piloting informed relevant modifications to the wording and content of the questionnaires to ensure appropriate content and length. Table 1 shows the research instruments used in this study.

Table 1. STARS Research Instruments and Contents

Instrument	Modules	Administration Mode
Arrival Survey	Background information and teacher roster.	In-person interview
Classroom Observation	Classroom processes and practices, teacher-child interaction, and student behavior.	Observation
Roll Call Tool	Teacher and pupil information, pupil grouping (by class or level), and pupil-level attendance.	Roll call
Pupil Counting Tool	Count of pupils in levels and class	Roll call
Learning Progress Tool	Pupil information, pupil levels in English and mathematics	Documentary review
Teacher Survey	Background characteristics, teacher supervision, support, teacher satisfaction, work stress/burnout, and implementation of targeted instruction pedagogy.	In-person interview

[8] The total teacher sample was larger than 630 (210*3) as some schools were multi-stream, i.e., a single school has multiple sections of a grade.

[9] To be eligible for this study, students must be present on the day of enumeration. For each grade, we randomly selected (a) a stream (if a multi-stream school) and (b) 15 students from each stream to participate in the study. If a stream or grade had fewer than 15 students, all students were included.

Instrument	Modules	Administration Mode
Head Teacher Survey	Background characteristics, school characteristics, and activities, participation in professional development activities, perceptions about their role, work stress/burnout, technology use, and implementation of targeted instruction pedagogy,	In-person interview
Circuit Supervisor Survey	Background characteristics; management and supervision; participation in professional development activities; perceptions about the role as a circuit supervisor; work stress and burnout; and technology use.	Phone or in-person interview
Student Survey	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);	In-person interview
Student English Assessment	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 Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) tools were included. Students were assessed on word recognition, reading a simple sentence, verb tense, oral and reading comprehension, and vocabulary.	Direct assessment
Student Mathematics Assessment	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. Students were assessed on number recognition, counting, addition, subtraction, word	Direct assessment

Instrument	Modules	Administration Mode
	problems, multiplication, division, simple fractions, and telling time.	

3.5 ETHICAL CONSIDERATIONS

IPA Institutional Review Board approved the protocol and questionnaires. Verbal consent was obtained for each respondent participating. All respondents were informed of the voluntary nature of participation and the confidentiality and anonymity of information. Each participating teacher received GH₵ 5 airtime while the head teachers and circuit supervisors received GH₵ 10 each (approximately US$1 and US$2, respectively).

3.6 DATA ANALYTICAL PROCEDURE

3.6.1 RESEARCH HYPOTHESES

Our design allowed us to test 5 primary hypotheses, two about the impact of any treatment relative to control (T1 vs. T2, T3) and three about the overall and marginal effect of the enhanced management treatment (T1, T2 vs. T3).

H1. Targeted instruction improves **student-level outcomes** relative to the control group (T2 > T1; T3 > T1). As targeted instruction focuses on the learning level and not the grade level of students, this pedagogy should improve test scores and student attendance.

H2. Targeted instruction improves **teacher-level outcomes** relative to the control group (T2 > T1; T3 > T1). As targeted instruction focuses on pedagogy, we expect it to improve teaching practices and student-teacher interactions. High-quality classroom-level outcomes include improved teacher attendance at school, teacher presence in the classroom, engagement with students, and use of learning materials.

H3. Management training for circuit supervisors and head teachers improves **teacher and school-level outcomes,** including teacher interactions with their managers, and quality of managerial practices, relative to the TI without enhanced management treatment and relative to control (T3 > T2, T3 > T1). As management training involves head teachers and circuit supervisors, we expect it to improve interactions among these personnel.

H4. Enhanced management training improves **classroom-level outcomes** relative to the TI without enhanced management treatment (T3 > T2). One barrier to implementation in Duflo, Kiessel, and Lucas (2018) was that teachers reported not feeling supported and

they felt this was one of many different things they were asked to do. Better-supported teachers would implement the project with fidelity, resulting in higher quality classroom-level outcomes, including improved teacher attendance at school, teacher presence in the classroom, engagement with students, and use of learning materials.

H5. Enhanced management training for circuit supervisors and head teachers increases student-level outcomes relative to the TI without enhanced management treatment (T3 >T2). This hypothesis depends on whether the enhanced management training caused differentially higher improvement in classroom and teacher level outcomes than targeted instruction alone.

3.6.2 IMPACT ESTIMATES

The estimation equation for this randomized controlled trial is shown below.

$$Y_{isj} = \beta_0 + \beta_1 TI_s + \beta_2 TI\_Mgmt_s + \delta' X_{isj} + \gamma_j + \epsilon_{isj} \qquad (1)$$

where *Y* is the outcome of interest for respondent *i* in school *s* in district *j*. *TI and TI_Mgmt* are dummy variables indicating the randomly assigned treatment status of the school. These indicators are mutually exclusive. Further, along with the omitted category of control, they are completely exhaustive. *X* is a vector of control characteristics, including the baseline value of the outcome variable if measured, and the round of the survey (if the question appears in multiple rounds). γ is a district fixed effect, our stratification cell. All standard errors were clustered at the school level.

Tests of H1 and H2 correspond to whether β_1 and β_2 are statistically different from 0.

Tests of H3 through H5 correspond to whether β_1 and β_2 are statistically different from each other.

3.6.3 MEASURES

We measured outcomes at the student, classroom, teacher, head teacher, and circuit supervisor levels. The primary outcome variables for students were attendance, standardized test scores on the mathematics exam at follow-up, standardized test scores on the English exam at follow-up, and standardized combined mathematics and English test score at follow-up. Subject-specific test scores were constructed using item response theory and standardized based on the baseline mean and standard deviation. We also tested the effect of the intervention on the combined test score. Control variables for the student-level regression were strata (district),

standardized baseline test scores, gender, age, and grade as well as a dummy for survey round (for the regression using attendance as the outcome variable).

Our bespoke test was able to capture latent student ability across the test score distribution. Despite extensive piloting, at the baseline we had a small mass of students scoring 0 on the English exam (7.7 percent) and about half that (2.8 percent) scoring the maximum. In math, both of these numbers were less than 1 percent. We adjusted the tests for follow-up, maintaining anchoring questions that allowed us to compare test scores across the two rounds. At follow-up, less than one percent of students scored the maximum or minimum on either of the tests. In the appendix, we provide the graph of the test information function that further shows our ability to measure latent ability throughout the test score distribution.

4. EVALUATION FINDINGS

4.1 DESCRIPTIVE STATISTICS

Table 2 presents the descriptive statistics. The sample used to construct the average is at the top of each column.

Table 2. Descriptive Statistics by Treatment Group

	All	Control	TI-Only	TI +Management
	(1)	(2)	(3)	(4)
Panel A: Students				
Male	0.47	0.47	0.46	0.47
Age	12.1	12.2	12.1	12.1
Baseline Combined Test Score	0.00	0.00	-0.01	0.01
Endline Combined Test Score	0.40	0.33	0.43	0.45
Panel B: Teachers				
Male	0.74	0.74	0.75	0.73
Age	31.4	31.6	31.6	31.0
Panel C: Head Teachers				
Male	0.82	0.78	0.86	0.83
Age	42.1	42.8	40.6	42.9
Panel D: Circuit Supervisors				
Male	0.91	--	0.91	0.90
Age	45.5	--	46.3	44.6

About half of the students were male. The average age of our students at baseline was just over 12. Recall that these were grade 4 and 5 students at baseline, surveyed in the final term of the year. Had these students started grade 1 on time at age 6 and progressed on pace, they would have been 9 to 11 at the time of the survey. Therefore, the average age of over 12 shows that many students started after age 6 and/or repeated a grade.

As we standardized our test scores based on the baseline mean and standard deviation, the average test score in the baseline was 0. The final row of Panel A shows a priori evidence of the success of the program—at the follow-up student test scores in two treatment arms were about 0.10 standard deviations (SD) larger than those in the control schools.

The remaining panels contain the summary statistics for the adult respondents. About three quarters of teachers were male with an average age of 31. Head teachers were even more likely to be male—82 percent—and were also older—42 years old. Almost all circuit supervisors were male (91 percent) and they were on average 46 years old.

4.2 STUDENT OUTCOMES

Table 3 contains the effect of the two interventions on standardized student test scores. Each intervention increased student's combined math and English score by about 0.11 SD (Column 1). When considering the math score alone, the estimated test score increase is about 0.13 SD (Column 2). We estimate that the English test scores increased by about 0.07 SD (Column 3). Even though the point values associated with the two interventions are somewhat different, based on an F-test of their equivalence, they are statistically the same (see "PValue Same Effect" for Columns 1 - 3 in Table 3). Therefore, we fail to reject that the two versions of the intervention had the same effect on test scores.

On average, control school test scores increased by 0.33 SD between the baseline and the follow-up, approximately one year. Therefore, our students learned an extra 33 percent over this same period.[10]

[10] Lucas et al. (2014) found literacy increases of about 25 percent of the control group change when evaluating a 1.5-year teacher training and learning materials intervention in Uganda. Lucas et al. (2019) found increase in test scores of 18 to 34 percent of a year of schooling after students were exposed to TCAI in Ghana for two years starting at the end of grade 1.

Table 3. Student Outcomes

	Combined Score	Math Score	English Score	Present at Spot Check
	(1)	(2)	(3)	(4)
TI only	0.108***	0.140***	0.065***	-0.031**
	(0.021)	(0.026)	(0.022)	(0.014)
TI+Management	0.107***	0.131***	0.076***	0.000
	(0.024)	(0.029)	(0.024)	(0.013)
Observations	5,608	5,608	5,608	11,569
PValue Same Effect	0.95	0.75	0.63	0.02
PValue Joint Zero	0.00	0.00	0.00	0.04
Control Group Mean	0.33	0.32	0.30	0.83
Notes: Sample: all students who took the endline assessments. Outcome: standardized test scores. All regressions include baseline controls for student grade, baseline math and English score, age, age-squared, female, and strata (district) fixed effects. Standard errors in parentheses, clustered at the school level. * p<0.10, ** p<0.05, *** p<0.01				

These increases in test scores are likely not due to changed attendance patterns. In Column 4 we test for the effect of the intervention on the likelihood that students were present at one of the two unannounced spot checks. Students in the TI-only intervention were 3 percentage points less likely to be present than those in the control group. We find no statistically significant difference between the intervention that included the management training and the control group. We further reject that the effect sizes are the same across the two interventions.

One of the motivations for the research was to assist students who were behind their grade-level competencies. We test for heterogeneity in effect by baseline test score by interacting baseline test score with each of the treatment variables and including these additional interaction terms as regressors. We find small, statistically insignificant, approximately 0 (0.01 and 0.002) coefficient estimates. Recall that all students were taught in a more homogenous setting, not just those who were behind grade-level. Therefore, while the intervention increased test scores, it did not narrow the within grade-level test score gap.

Further, teachers of different levels of experience might have found the program easier or harder to implement. This analysis is slightly complicated because, by design, students in treatment schools might have worked with one teacher during non-TI hours and another teacher during TI-hours, their TI-math teacher might have been different than their TI-English teacher,

and their TI-teachers may have changed during the course of the implementation as students changed levels. Therefore, to estimate heterogeneity by teacher years of experience, we calculate the average years of teacher experience for the P4 through P6 teachers at baseline and interact this average with the two treatment indicators. As with the baseline test score interaction, we find small, statistically insignificant, approximately 0 (0.003 and -0.002) coefficient estimates on the interaction terms.

One concern with any test score changes is differential selection into the test based on who was present the day of the follow-up exam. Due to robust tracking methodologies, we were able to track about 96 percent of control group children. In Table 4 Column 1 we test for differential attrition by treatment status. Students in the management arm were 1.5 percentage points less likely to take the follow-up exam (Column 1). In Column 2 we test for differential attrition by treatment status and baseline test score and find no relationship between the interaction of treatment and student baseline test score and the likelihood we could track them for follow-up testing. As attrition differences between treatment arms are small and not systematically related to both test scores and treatment status, these differences are likely not biasing our results.[11]

Table 4. Student Participation in Follow-up Exam

	Participated in Follow-up	
	(1)	(2)
TI-only	-0.003	-0.003
	(0.006)	(0.006)
TI-only X Baseline Score		0.006
		(0.007)
TI+Management	-0.015**	-0.015**
	(0.007)	(0.007)
TI+Management X Baseline Score		-0.002
		(0.008)
Observations	5,893	5,893
PValue Same Effect	0.06	
PValue Joint Zero	0.05	
Control Group Mean	0.96	

[11] The Lee (2009) bounds of the coefficient on TI+Management in the previous table are 0.098 and 0.123.

4.3 TEACHER AND CLASSROOM IMPLEMENTATION OUTCOMES

In the first three columns of Table 5, we test for the effect of the interventions on teacher behaviors. Teachers were no more (or less) likely to be present when the enumeration team arrived (Column 1). Teachers were more likely to be in the classroom—13 percentage points in TI-only and 11 percentage points in TI+Management, statistically indistinguishable from each other (Column 2). Teachers in both treatments were also more likely to be using STARS teaching and learning materials—29 percentage points in TI-only and 25 percentage points in TI+Management.

Table 5. Teacher Attendance and Classroom Activities at Unannounced Spot Checks

	Teacher Present At Arrival	Teacher in Classroom	Any TLM STARS use	Students in Levels (Enumerator Report)
	(1)	(2)	(3)	(4)
TI-only	0.037	0.131***	0.287***	0.577***
	(0.035)	(0.032)	(0.018)	(0.048)
TI+Management	0.021	0.113***	0.254***	0.621***
	(0.039)	(0.034)	(0.018)	(0.046)
Observations	1,456	2,462	2,423	411
R-squared	0.08	0.07	0.15	0.45
PValue Same Effect	0.70	0.54	0.12	0.44
PValue Joint Zero	0.57	0.00	0.00	0.00
Control Group Mean	0.63	0.68	0.01	0.00

Notes: Regression includes observations from all three follow-up rounds. All regressions include district and round fixed effects, with standard errors clustered at the school level. Standard errors in parentheses, clustered at the school level. * p<0.10, ** p<0.05, *** p<0.01

In the final column of Table 5, we test at the school level whether students were correctly divided by learning levels instead of grade levels. About 58 percent of the time students were divided by learning levels in the TI-only intervention and 62 percent of the time in the TI+Management intervention (statistically indistinguishable from each other).

Observing students divided by learning level instead of grade is only one possible measure of TI implementation. We find that across both treatments about 90 percent of schools reported that they had implemented TI at least once in the past week, had conducted a leveling exam at either the start of the current term or the end of the previous term, and were able to show the enumerator the leveling exam score sheet. As with observing splitting students by learning level, these effects are not different across the two treatment interventions.

4.4 HEAD TEACHER AND CIRCUIT SUPERVISOR OUTCOMES

For head teachers' outcomes, we directly observed head teachers and asked teachers to report on their interactions with their head teacher.

Columns 1 through 3 of Table 6 displays the head teacher outcomes. Both interventions increased the likelihood that the head teacher was present when we arrived by 12 (TI-only) to 16 (TI+Management) percentage points. On average in the control group schools, head teachers were present only 42 percent of the time. Therefore, this represents an increase of 28 to 38 percent.

Table 6. Head Teachers and Circuit Supervisor

	HT Present	# Times HT Observes Class (30 minutes)	Any HT Feedback (1/0)	# Times CS Observes Class (30 minutes)	Any CS Feedback (1/0)
	(1)	(2)	(3)	(4)	(5)
TI-only	0.117*	0.683***	0.117***	0.417***	0.126***
	(0.063)	(0.148)	(0.030)	(0.068)	(0.039)
TI+Management	0.158**	0.855***	0.143***	0.617***	0.172***
	(0.065)	(0.127)	(0.031)	(0.077)	(0.038)
Observations	417	1,288	1,809	1,415	1,213
PValue Same Effect	0.53	0.24	0.36	0.01	0.08
Mean Dep in Control	0.42	0.74	0.69	0.00	0.00

Notes: Regressions include responses from all three follow-up rounds. All regressions include district and round fixed effects, with standard errors clustered at the school level. Columns 1, 3, and 5: Linear probability models. Columns 2, 5, and 6: These questions were not asked during the first observational study. Standard errors in parentheses, clustered at the school level. * p<0.10, ** p<0.05, *** p<0.01

When we asked teachers the number of times that the head teacher observed their teaching for at least 30 minutes, teachers in the TI-only treatment reported 0.68 more observations than the control group and those in the TI+Management treatment 0.86 more observations than the control group (Column 2). Further, head teachers were more likely to provide teachers feedback (Column 3)—12 percentage points more in the TI-only and 14 percentage points more in the TI+Management arms—and teachers were more likely to think that this feedback was useful—

about 12 percentage points in both arms. In all cases, we fail to reject that these effects are equal across the two treatment interventions.

About 57 percent of head teachers were able to produce the TI handbook to show the enumerator with equivalent rates across the two treatment arms.

The final two columns of Table 6 are based on teacher reports about circuit supervisors. In control schools, circuit supervisors did not conduct any classroom observations of at least 30 minutes. In contrast, the TI-only intervention increased that number by 0.42 visits and the TI+Management intervention increased that number by 0.62 visits. Unlike all previous measures presented, we reject the statistical equivalence across the interventions. Circuit supervisors were also more likely to provide feedback to teachers—13 percentage points in TI-only and 17 percentage points in TI+Management. As with the number of observations, we reject that these two numbers are statistically equivalent. Therefore, the additional management training that the circuit supervisors received appears to have increased both the number of observations they conducted and the likelihood that they provided teachers feedback beyond any changes from the school being a TI school.

5. COST-EFFECTIVENESS ANALYSIS OF THE STARS PROJECT

The cost-effectiveness analysis (CEA) of the STARS program was done from both a societal and a budgetary perspective. To compute the CEA, we used two pieces of data: an estimate of the STARS program's impact and the cost of the program. To calculate the cost of the program (Appendix Tables 4 and 5), we used UNICEF's record of transaction details on the direct expenses incurred in implementing the targeted instruction and enhanced management program of the STARS program. We also included the opportunity costs of the program by including the results of a telephone survey on participants' time spent working on the STARS program.

Based on JPAL's methodology of computing program cost for CEA, the program's cost comprised of (a) program administration and staff costs, (b) implementation and program material costs, (c) transportation and per diems, (d) targeting/visibility costs, (e) training and (f) office costs. The program's cost was determined based on the following assumptions (Dhaliwal et al., 2013): (a) inflation is calculated using GDP deflators; (b) average inflation from the base year to the year of analysis was computed by assuming that the program costs were incurred on the first day of each year; (c) costs were expressed in terms of 2018 USD, with local currencies exchanged using standard exchange rates; and (d) a ten percent discount rate applied for costs incurred over multiple years.

There was a total cost of $243,585.42 to implement the targeted instruction program with a total cost per student of $29.22 during the 2018/2019 school year. The total cost of implementing the targeted instruction and management training program was $399,916.42 with $52.62 as the cost per student during the 2018/2019 academic year. Comparison of the program cost estimates with the program's estimated impact gives measures of the cost-effectiveness of the STARS program to compare with alternative means of obtaining the same benefits. With an increase in math and English achievement of 0.108 standard deviations (SD) for the targeted instruction arm, the cost per child is $29.22 and the cost per additional SD increase in math and English achievement is $270.56. Using Evan and Yuan (2018)'s calculations of a metric to demonstrate the potential returns to learning interventions, we can estimate that the targeted instruction arm of the STARS program results in 1.74 - 2.51 additional years of schooling per $100 spent. This estimate is intended to be demonstrative rather than predictive, as we are assuming a linear relationship between spending and impact and that interventions across countries have the same effect. Based on our estimates of the targeted instruction and enhanced management program [0.107 SD increase in math and English achievement], real annual cost of increasing math and English achievement score for a student by one SD was $491.78. To rephrase this, the targeted instruction and enhanced management arm of the STARS program results in 0.96 - 1.38 additional years of schooling per $100 spent.

The cost-effectiveness ratios for increasing math and English achievement scores through the implementation of targeted instruction was more favorable, compared to the cost-effectiveness ratios for achieving the same outcome through the implementation of targeted instruction and enhanced management training, because of the observed substantially lower program cost.

6. CONCLUSIONS

Building off prior work in Ghana and India we tested the effect of teacher-led targeted instruction, both with and without additional training for managers, on student test scores and attendance, and teacher, head teacher, and circuit supervisor activities. We find that both interventions increased test scores on both English and Math by about 11 percent of a SD on the combined test score. This was not due to increased student attendance as students were less likely to be present in the TI-only schools relative to the control schools.

Teachers were no more likely to be present at school, but they were more likely to be in the classroom and using STARS teaching and learning materials. Students were divided by learning level instead of grade level about 60 percent of the time. Head teachers were more likely to be present, performed more classroom observations of at least 30 minutes, and provide feedback.

While many of the student, teacher, and head teacher point values were larger for the intervention that included management training, the two interventions were statistically indistinguishable from each other.

Both interventions increased the number of classroom visits of at least 30 minutes that circuit supervisors performed and increased the likelihood that the circuit supervisors provided feedback to the teaching. For these outcomes, the point values were larger for the intervention that included the management training and we reject the statistical equivalence across the two interventions.

We present below our findings in relations to the stated research objectives:

The first evaluation objective relates to whether targeted instruction improves student test scores in upper primary grades. This study shows that this program increased test scores in math by 0.13-0.14 SD and in English by 0.07-0.08 SD. We are conducting a follow-up during this academic year to test whether these impacts are sustained.

Our second research objective sought to establish whether monitoring, managing, and coaching performed by existing circuit supervisors and head teachers increased the likelihood that teachers implemented targeted instruction. We find that both treatment groups were equally likely to implement targeted instruction. Both interventions increased teachers' likelihood to be present in class, use STARS teaching and learning materials, correctly group students by learning levels, implement targeted instruction lessons at least once in the past week, and conduct a leveling exam at either the start of the current term or the end of the previous term. We cannot reject that the results are equal in both treatment arms.

Our findings also suggest that the targeted instruction intervention also led head teachers and circuit supervisors to increase classroom observations and provide feedback. While head teachers in the enhanced management arm reported 0.16 more classroom observations, this effect is not statistically different from zero. The primary effect of the management intervention is from circuit supervisor reports. Circuit supervisors reported 0.20 more classroom observations and increased the likelihood that they provided feedback to teachers by 0.05 percentage points. These differences did not result in different levels of implementation of targeted instruction.

Our third research objective sought to determine whether enhanced monitoring, managing, and coaching lead to higher learning gains in students. While targeted instruction increased student learning overall, our findings demonstrate that there were no differential effects on student learning outcomes when targeted instruction was paired with enhanced management. The current analysis leaves this as an open question. We are currently exploring these results through additional data analysis and qualitative interviews.

Relative to the TCAI findings, the achievement point values are larger than those estimated for students who were subject to teacher-led targeted instruction for a little more than two years, starting in grade 1 and ending in grade 3. They also represent a larger portion of a year of schooling. The rate at which students were observed divided by learning levels instead of grade levels was also substantially higher in STARS, even in the treatment without additional management support. One important difference between the two studies is the level of direct government involvement. During STARS, national level monitors visited many of the schools and all of the districts. Their involvement made it clear to these schools (and their district level officials) the importance that the national level officials put on this implementation. Even though national stakeholders were supportive of TCAI, their enthusiasm likely did not permeate to individual schools.

REFERENCES

Adhvaryu, A., Kala, N., & Nyshadham, A. (2018). The skills to pay the bills: Returns to on-the-job soft skills training. NBER Working Paper No. W24313, 2018.

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.

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.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. (2017). From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. The Journal of Economic Perspectives, 31(4):73-102.

Bloom, N., Schankerman, M., & van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. Econometrica, 81, 1347-1393.

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

Dhaliwal, I., Duflo, E., Glennerster, R., Tulloch, C. (2012). Comparative cost-effectiveness analysis to inform policy in developing countries. https://www.povertyactionlab.org/sites/default/files/publications/CEA%20in%20Education%202013.01.29_0.pdf

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas (2020). External Validity: Four Models to Increase Student Learning. Working Paper

Evans, David K. and Fei Yuan. 2018. "Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms." World Bank Policy Research Working Paper 8752. https://openknowledge.worldbank.org/bitstream/handle/10986/31315/WPS8752.pdf?sequence=1&isAllowed=y

Lucas, Adrienne M., Patrick J. McEwan, Moses Ngware, and Moses Oketch (2014). "Improving early-grade literacy in East Africa: Experimental evidence from Kenya and Uganda." *Journal of Policy Analysis and Management* 33, no. 4: 950-976.

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.

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.

APPENDICES

Appendix A

FIELD STAFF RECRUITMENT AND TRAINING

All data collectors were carefully selected and receive relevant and sufficient specialized training and ongoing support to facilitate their participation in the data collection process

RECRUITMENT

Field staff were recruited based on IPA's short-term recruitment policy, which followed a competitive process. The recruitment of field staff for each wave of data collection was based on the following selection criteria: a minimum of a diploma certificate in any related field; at least a one-year experience in school-based data collection; experience in the use of computer assisted personal interviewing (CAPI); for team leaders and supervisors: experience managing fieldwork teams; and working knowledge about the 20 UNICEF-supported districts. Thirty percent of the field staff was shortlisted for training to account for attrition and ensure the hiring of qualified candidates from the pool of trainees.

CLASSROOM TRAINING

Classroom training was done based on the roles of field staff in the STARS evaluation activities. It involved training different field staff, namely, field supervisors, team leaders, and interviewers. Classroom training comprised presentations, questions and answers, group discussions, and role-plays. The presentations centered on information on the STARS project, questionnaire content and review, IPA and survey protocols, and use of CAPI for data collection. The training for each wave of data collection was non-residential. The Research Associate and Field Manager led the training. Trainees' performance was evaluated to gauge their progress; provide performance feedback to both the trainers and the trainees; as well as to select the best candidates for each wave of data collection. Trainees' performance was assessed using quizzes, role-plays, and field practices or school visits.

Appendix Table 1. STARS Evaluation Training Status

Data Collection Wave	Period	Trainee Attendance		
		Field supervisors	Team leaders	Interviewers
Baseline Survey	18th - 23rd May 2018	12	15	56
Observational Study I	10th - 17th October 2018	N/A	4	28
Observational Study II	4th - 8th February 2019	N/A	4	28
Follow-up Survey	21st - 29th May 2019	4	10	40

FIELD PRACTICE

Field practice was conducted as part of the training in seven implementation pilot schools in Asikuma-Odoben-Brakwa District to help the trainees apply what they learned during the training, and for the trainers to provide the trainees with specific and constructive feedback during a debriefing session. The field practice was conducted on 16th October 2018. Twenty-six field staff participated in the field practice. The field practice involved observing 7 classrooms and interviewing 21 teachers and 7 head teachers ().

Appendix Table 2. STARS Evaluation Field Practice

Data Collection Wave	Date	Sample Coverage				
		Schools	Head Teachers	Teachers	Children	Circuit Supervisors
Baseline Survey	23rd May 2018	10	10	30	300	5
Observational Study I	16th October 2018	7	7	21	N/A	3
Observational Study II	7th February 2019	6	6	18	N/A	2
Follow-up Survey	27th May 2019	7	7	21	210	3

FIELD TEAM STRUCTURE AND COMPOSITION

The composition and field team structure used for the STARS evaluation activities depended on the wave of data collection. As shown in Appendix Table 3, the average field staff used for the baseline and follow-up surveys was 54 field staff. Field teams were varied based on the specific data collection activity.

Appendix Table 3. Field Team Structure and Composition

Data Collection Wave	Number of Field Teams	Total Field Staff
Baseline Survey	12	64
Observational Study I	4	22
Observational Study II	4	22
Follow-up Survey	8	43

EVALUATION TIMELINESS

Appendix Table 4 presents the timeliness for conducting the data collection activities. Prior to field staff training, the research instruments were piloted to ensure that the instruments were applicable within the Ghanaian context.

Appendix Table 4. STARS Evaluation Timeliness

Activity	Piloting	Field Staff Training	Survey Period
Baseline Survey	2nd - 8th May 2018	18th - 23rd May 2018	31st May 2018 - 23rd June 2018
Observational Study I	19th - 20th September 2018	10th - 17th October 2018	25th October - 30th November 2018
Observational Study II	21st January 2019	4th - 8th February 2019	18th February - 19th March 2019
Follow-up Survey	13th - 14th May 2019	21st - 29th May 2019	4th June - 26th July 2019

DATA-COLLECTION PROCEDURES

Data were collected using CAPI based on the SurveyCTO platform and Samsung tablets, incorporating IPA's data management system. IPA Ghana's research quality team programed the questionnaires. The programmed instruments included constraints, skip patterns, relevant commands to automate the administration process and automatically check inconsistencies or errors associated with the administration of the instruments on the field. The CAPI application was bench-tested during the training field practice. Modifications of the questionnaires based on the pilot and field practice were incorporated into the electronic versions of the questionnaires. All survey instruments were administered in English. Data were collected through direct observations of classrooms and interviews conducted in-person and/or via phone.

This study has four respondent types: circuit supervisors, head teachers, teachers, and students. Each respondent type completed a different survey. Circuit supervisors completed a survey over the phone that covered a range of topics from their personal background to their circuit management practices. Head teachers' surveys were conducted in-person and covered a range of topics from information on their school and characteristics, to their personal background and school management practices. Teacher surveys were conducted in-person and covered a range of topics from information on their personal background to the support that they received from their managers.

Students completed two instruments: a short demographic survey and a student assessment on their math and English skills. We based the student assessment on previous national, international, and study tests. The majority of questions came from the examinations used in Duflo, Kiessel, and Lucas (2019). Those tests were developed by education stakeholders in the Ministry of Education to reflect grade 1-3 material. Based on the piloting and findings from that study, many students in P4 and P5 still tested at that level. We added additional questions that were inspired by questions from the National Education Assessment P3 and P6 exams. Finally, we added questions inspired by the Ghana versions of the EGRA and EGMA. Trained enumerators conducted the student assessments one-on-one. Enumerators either read the question aloud or showed questions to students to read, depending on the specific instructions for each question. Tests were semi-adaptive: all students started with the same questions. Those who performed poorly on these answered a set of easier questions. Students who performed well on the early questions answered a set of harder questions. Itemized response theory was used to convert each students' score to a common scale.

DATA QUALITY SYSTEMS AND DATA MANAGEMENT

Data collection activities were monitored to assess the (a) performance of the fieldwork teams in administering the various instruments and (b) quality of the data being collected. Field teams were monitored using IPA's standardized monitoring tool, hosted on SurveyCTO. On average, at least two different monitors monitored each field staff during the data collection period. The results from the monitoring largely showed the field staff strictly adherence to the established protocols. This was partly due to the use of experienced enumerators, feedback-based training, and the provision of timely feedback to the fieldwork team.

Apart from observing the interviews or observations, a standardized and coordinated system of checks and systems were developed and implemented in managing the data flow, collection, cleaning, and storage. These coordinated systems helped to ensure the accessibility, quality, reliability, and timeliness of the data. Using the IPA's Data Management System, high-frequency

checks were run daily to identify inconsistencies, electronic programming errors, and enumerator errors. The high-frequency checks indicated a minimal violation of the data quality checks such as duplicate IDs, missing values, constraints, skip patterns and survey logic or inconsistencies. Also, 10% each of the completed Head Teacher Survey, Teacher Survey, and Circuit Supervisor Survey were audited to establish whether there were variations in key outcome variables. The audit checks showed that discrepancies were largely within the acceptable range. During data collection and following the completion of fieldwork, data were edited and cleaned using STATA do-files. Data were encrypted using BoxCryptor from the point of collection to storage.

Appendix B

Appendix Table 5. Cost Summary of Targeted Instruction

Cost Ingredients	Total Cost/Yr Local Currency	Currency	Year Incurred	Total Cost/Yr, USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
Program administration and staff costs - Year 1	GHS 49,664.06	GHS	2018	$10,831.09	$10,831.09	$10,831.09	$10,831.09
Implementation and program material costs - Year 1	GHS 211,302.00	GHS	2018	$46,082.23	$46,082.23	$46,082.23	$46,082.23
Transportation and per diems - Year 1	GHS 1,236,275.88	GHS	2018	$269,615.76	$269,615.76	$269,615.76	$269,615.76
Targeting / visibility costs - Year 1	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Training - Year 1	GHS 736,594.70	GHS	2018	$160,641.76	$160,641.76	$160,641.76	$160,641.76
Office Costs - Year 1	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Program administration and staff costs - Year 2	GHS 25,996.00	GHS	2018	$5,669.39	$5,669.39	$5,669.39	$5,669.39
Implementation and program material costs - Year 2	GHS 8,900.00	GHS	2018	$1,940.97	$1,940.97	$1,940.97	$1,940.97
Transportation and per diems - Year 2	GHS 221,003.62	GHS	2018	$48,198.03	$48,198.03	$48,198.03	$48,198.03
Targeting / visibility costs - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Training - Year 2	GHS 798,822.84	GHS	2018	$174,212.92	$174,212.92	$174,212.92	$174,212.92
Office Costs - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Time Costs - Teachers	GHS 3,334,170.41	GHS	2018	$727,139.39	$727,139.39	$727,139.39	$727,139.39

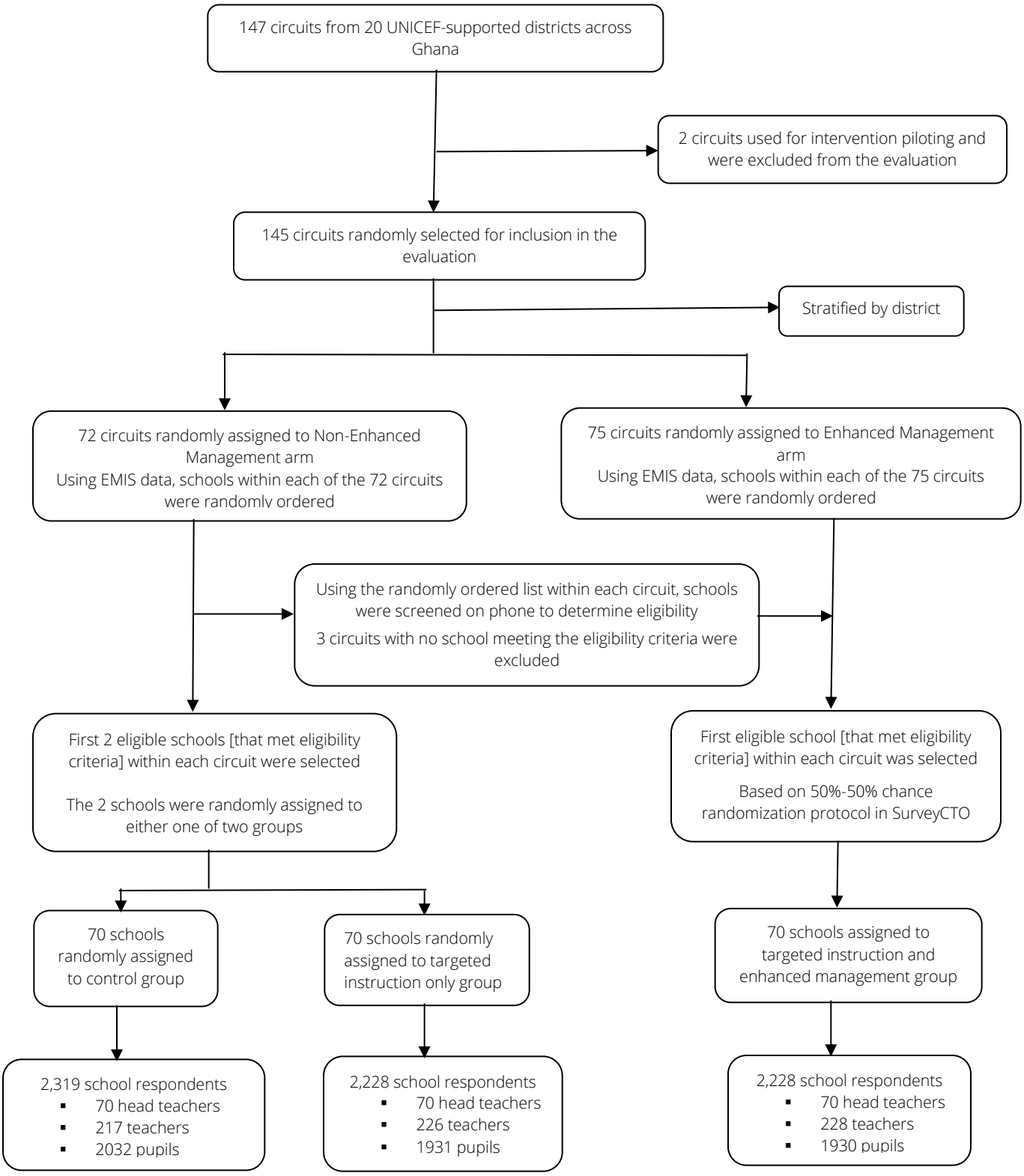
Time Costs - Head Teachers	GHS 930,384.15	GHS	2018	$202,904.73	$202,904.73	$202,904.73	$202,904.73
Time Costs - Circuit Supervisors	GHS 11,481.79	GHS	2018	$2,504.03	$2,504.03	$2,504.03	$2,504.03
Total Cost (without opportunity costs)							$243,585.42
Total Cost per Child (without opportunity costs)							$29.22
Total Cost (with opportunity costs)							$1,649,740.30
Total Cost per Child (with opportunity costs)							$197.91

Appendix Table 6. Cost Summary - Training in Targeted Instruction + Management Training

Cost Ingredients	Total Cost/Yr Local Currency	Currency	Year Incurred	Total Cost/Yr, USD	Total Cost/Yr, Base Year USD	PV of Cost Stream, Base Yr USD	Total Cost, Yr of Analysis USD
Program administration and staff costs - Year 1	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Implementation and program material costs - Year 1	GHS 140,152.72	GHS	2018	$30,565.49	$30,565.49	$30,565.49	$30,565.49
Transportation and per diems - Year 1	GHS 74,253.46	GHS	2018	$16,193.72	$16,193.72	$16,193.72	$16,193.72
Targeting / visibility costs - Year 1	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Training - Year 1	GHS 500,962.24	GHS	2018	$109,253.38	$109,253.38	$109,253.38	$109,253.38
Office Costs - Year 1	GHS 1,460.00	GHS	2018	$318.41	$318.41	$318.41	$318.41
Program administration and staff costs - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Implementation and program material costs - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Transportation and per diems - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00

Targeting / visibility costs - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Training - Year 2	GHS 340,250.22	GHS	2018	$74,204.17	$74,204.17	$74,204.17	$74,204.17
Office Costs - Year 2	GHS -	GHS	2018	$0.00	$0.00	$0.00	$0.00
Time Costs - Teachers	GHS 3,219,081.64	GHS	2018	$702,040.02	$702,040.02	$702,040.02	$702,040.02
Time Costs - Head Teachers	GHS 1,070,769.51	GHS	2018	$233,520.96	$233,520.96	$233,520.96	$233,520.96
Time Costs - Circuit Supervisors	GHS 1,186,281.25	GHS	2018	$258,712.58	$258,712.58	$258,712.58	$258,712.58
Total Cost (without opportunity costs)							$399,916.42
Total Cost per Child (without opportunity costs)							$52.62
Total Cost (with opportunity costs)							$1,424,808.72
Total Cost per Child (with opportunity costs							$187.47

Appendix C. Flow Chart of Sample Selection



Appendix D. Test Information Function