Every Child Counts: Adapting and Evaluating Targeted Instruction Approaches into a New Context through a Nationwide Randomized Experiment in Ghana
(DRAFT Working Paper)

Annie Duflo and Jessica Kiessel, 2015

Through a large-scale, nationwide experiment, this study evaluates four interventions that adapt evidence-based programs from India and Kenya into the Ghanaian context. Three of these interventions involve using assistants or training teachers to target instruction at the child’s level and teach low performing students the basic skills to that they don’t get left behind, and one intervention isolates the effects of adding an assistant alone. The results confirm that targeted instruction, despite challenges inherent to implementing within a government system, leads to improvements in learning levels: Two interventions using teacher assistants to provide remedial classes to the lowest third of the class, either during or after school hours, led to improvements in basic skills for Grade 2 and 3 students who had been in the program for two years or more (0.09 to 0.1 standard deviations), and to larger improvements and on a larger set of skills for Grade 3 students (0.14 to 0.18 standard deviations). These effects persist, as results for Grade 4 students who have been out of the program for one year are almost similar to those of Grade 3 students. Teacher-led targeted instruction, without the support of assistants, also led to improvements in basic skills (0.11 standard deviations), but had a smaller effect (0.1) on grade 3 students’ overall test scores, which is not surprising because they were less likely to enforce the program than the assistants. Results of an intervention involving adding assistants, but without the targeted instruction component, confirm that targeted instruction was a critical component of the approach, since results are for the most part insignificant. This study also investigates how implementation challenges affected the program’s impacts, as well as the importance of considering the school environment when determining the best implementation approach. It also highlights the importance, for policy relevance, of evaluating similar concepts across contexts, while adapting them to the new environment, and at scale.
1. Introduction

Under the ambitions of the Millennium Development Goals, tremendous progress has been made in the last fifteen years to achieve universal enrollment in primary school. However, learning levels have not matched this progress and many children are in school but not learning, especially in low and middle income countries. According to Ghana’s 2013 National Education Assessment, only 28.1% of grade 3 pupils reached national proficiency levels in English and 22.1% reach proficiency in math. While in 2010 nationwide survey Ghana showed, we found that only 6% of grade 3 students can read a simple grade 1 level text. While more needs to be learned about improving school quality and increasing overall learning levels, a number of interventions have been rigorously evaluated so as to better understand methods of improving student learning levels. Through these evaluations, one lesson, that has emerged clearly is that targeting instruction at the child’s level is an effective method of improving basic literacy and numeracy skills.

A number of programs aimed at improving education quality has focused on the large number of untrained teachers; and the necessity of providing professional training and improving teachers’ deployment. These are important goals. However, long-term teacher training and adequate teachers’ deployment require long-term investments that, given budget and management constraints faced in many countries, may not be quickly achievable. As each year a cohort of students move through the system, there is an urgent need for quality issues to be addressed with measures that can have an impact quickly with a limited budget.

Other interventions have focused on the resource gap in primary schools in the developing world. However, a number of studies have shown that improving standard resources alone—such as providing more of the same textbooks, flip charts, or additional teachers — may not achieve the desirable goals for a large fraction of the pupils, especially if they have not acquired the basic skills (Glewwe, Kremer, and Moulin 2002; Glewwe, Kremer, Moulin and Zitzewitz, 2004).

In contexts of large class sizes, high pupil-teacher ratios and the need for additional teachers has also been emphasized. Research suggests, however, that reducing the class size without any other changes may not be addressing the main issue. A study in Kenya demonstrated that adding additional teachers only produced meaningful change when classes were divided by initial learning levels. When tracked by ability level, pupils at all levels, high and low, benefited, and the differences persisted a year after the program had finished (Dupas, Duflo, and Kremer, 2008).

The movement towards free primary education for all has increased the enrolment rates of students from poorer backgrounds who are more likely to have illiterate parents and limited support from home. As a result, teachers often face crowded classes and pupils with very different levels. Disparities in learning levels and pressures to complete the standard curriculum, makes it challenging for teachers to target instructional time at the right level for all pupils, particularly for students lagging behind. If teachers focus primarily on the top of the class, it explains that reducing the class size does not affect learning levels, and that additional resources are more likely to only benefit the top of the class.

A study in two Indian cities demonstrated that having minimally-trained community tutors provide daily remedial classes focused on basic skills to the weakest students improved their basic skills in reading and
math. The ‘Balsakhi’ program, implemented by an Indian nongovernmental organization named Pratham, increased test scores of all children in treatment schools by an average of 0.14 standard deviations in the first year, and 0.28 in the second year of treatment, as compared to schools in the control group. The effects were mostly concentrated on the lowest performing students (those targeted by the program), whose learning levels were 0.6 standard deviations higher than in the control group. This evaluation indicated that targeting instruction to children’s ability levels can lead to significant increases in basic literacy and numeracy skills and at a lower cost relative to other educational programs. (Banerjee et. al, 2005)

The common theme in both the Kenya and India studies is that in the presence of heterogeneous learning-levels, learning levels can be improved by targeting instruction at the child’s level in classrooms. As mentioned, this can be achieved by dividing classes by ability level when there is more than one class per grade, or by having low-cost and less qualified tutors provide remedial cases during school hours. There are other ways to implement this concept. For example, an evaluation in India, also through Pratham, demonstrated that computer assisted learning, with the appropriate software, can be used effectively to adjust teaching and exercises to a child’s level (Banerjee et. al, 2005). Targeting instruction at the child’s level can also be achieved through summer camps targeted at the lower half of the class.

Given the similarities in educational context across Ghana, India and Kenya and the ready availability of untrained youth in Ghana, the evidence from these programs indicated an opportunity to replicate and extend the results in a new context. This paper presents the results of collaboration between the Ghana Education Services (GES), Innovations for Poverty Action (IPA), the Ghana National Association of Teachers (GNAT), and the National Youth Employment Program (NYEP), which strived to do just that by adapting these insights to the Ghanaian context through a pilot program named, the Teacher Community Assistant Initiative (TCAI).

The results confirm that targeted instruction, despite challenges inherent to implementing within a government system, can lead to improvements in learning levels: Two interventions using teacher assistants to provide remedial classes to the lowest third of the class, either during or after school hours, led to improvements in basic skills for Grade 2 and 3 students who had been in the program for two years or more (0.09 to 0.1 standard deviations), and to larger improvements and on a larger set of skills for Grade 3 students (0.14 to 0.18 standard deviations). Teacher-led targeted instruction, without the support of assistants, also led to improvements in basic skills (0.11 standard deviations), but had a smaller effect (0.1) on grade 3 students’ results on the overall test, which is not surprising because they were less likely to enforce the program than the assistants. Results of an intervention involving adding assistants, but without the targeted instruction component, confirm that it was a critical component of the approach, since results are for the most part insignificant. These effects persist, as results for Grade 4 students who have been out of the program for one year are almost similar to those of Grade 3 students. This study also highlights how implementation challenges affected the program’s impacts, as well as the importance of considering the school environment when determining the best implementation approach.

This paper is divided as follows: First we describe the Ghanaian educational context as well as the process of translating evidence-based insights from one context to another. Then we described the
evaluation design, the implementation, and the results. We end with reflections on policy implications and highlight future research needs.

2. The Context and the Interventions

2.1 The Ghanaian Education Context
Achieving universal education was a key priority at Ghana’s independence in 1957 and early on Ghana was known for investing in its education system. Even now, nearly 60 years later, the Ghanaian government is estimated to spend approximately 30% of its budget on education. However, nearly all of it goes towards payroll. While primary gross enrolment rates in Ghana have increased from 75.7% in 2002-2003 to 97.8% in 2011-2012, like in so many other countries in Africa, low achievement rates persist.

Currently, very few students in early primary gain the basic skills required to succeed in school. As mentioned earlier, the 2013 National Education Assessment (NEA) in Ghana showed that only about a quarter of students reach proficiency levels in English and Math, and 42.9% and 41.9% failed to reach even minimum competency in math and English respectively. NEA results also showed wide regional differences. For example, 57.1% of students in Greater Accra achieved proficiency in English and 43.9% in math on expected proficiency, while in the Northern region, 14% of students achieved proficiency in English and only 10% did so in math. The 2011 NEA also found wide gaps in achievement for multi-grade schools compared to non-multi-grade schools, in fact in 2011 only 5.2% achieved proficiency compared to 17.6% in non-multi-grade schools. Boys and urban schools had comparatively higher levels of achievement, although all were below the ideal.

Corroborating the above findings, in 2010, our baseline survey reveals that only 6% of grade 3 students could read a simple text at grade 3 level and about 18% of grade 3 students could not identify alphabets. Figure 1 shows the gap between students achievement based on these findings and what students should have achieved according to the national curriculum. Thus, clearly illustrating a growing gap between what students actually learn and what they should be learning. This also suggests that students who fail to learn basic skills in earlier grades fall further behind as they are promoted to new grade-levels. Therefore, ensuring students attain basic skills at an early grade is necessary for their longer term participation in the education system. In addition to growing gaps between expected and actual learning levels, there are gaps between levels within the classroom. Figure 2.1 and 2.3 highlight a large heterogeneity in learning levels. In such a context, it is understandably challenging for teachers to target instruction appropriately. In order to complete the school curriculum, teachers often need to leave a large fraction of pupils behind.

_____________________

1 The survey mentioned here is the Teacher Community Assistants initiative Baseline Survey in 500 schools.
2 The TCAI evaluation test covers the critical areas of the national syllabus. The red line shows what students in grades 1 to 3 scored on the TCAI baseline test and the green line shows what students in grades 1 to 3 of the TCAI control group sample scored at the second endline test. The blue line shows what students should have actually learnt based on the national curriculum.
2.2 Translating targeted instruction into the Ghanaian context

Ghana was a clear fit for a program designed to improve low learning levels through a targeted instruction model. Not only did Ghana need innovative and cost-effective solutions to improve learning levels of primary students, Ghana had number of similarities with Kenya and India, where targeted instruction had been successful. The country also faced large gaps in learning levels within primary classrooms, had a historically similar school systems and trained teachers to move systematically along a standard curriculum that was designed at a time when only elite students came to school, and therefore likely inappropriate to the current school population. Furthermore, the Ghanaian government had demonstrated willingness to learning gaps through the launch of a large National Literacy Acceleration Program (NALAP), which promoted the use of mother tongue in early grades.

Yet, important institutional and cultural differences also existed in the Ghanaian context that could have affected the impact of targeted instruction. As Ghana did not have a nationwide education non-profit,
like Pratham, any large-scale program would need to be implemented directly through the Ghanaian government education system.\textsuperscript{3} Volunteerism, important to Pratham’s Balsahki model, is uncommon in Ghana; and Ghana only trains the number of teachers they need. Therefore, a paid cadre of youth was needed to achieve national scale. The existence of the National Youth Employment Program (NYEP) that was managed Ministry of Youth and Sports at the time\textsuperscript{4} provided an opportunity to use an existing mechanism and financing stream to create such a cadre, although given a strong teachers’ union, the use of untrained teaching staff was a potential political risk.

In addition, community and school-level differences came up when piloting the intervention on the ground. For example, while it is common in India for children to sit on mats during pull-out classes, sitting on mats is often associated with Islam in Ghana, which made Ghanaian education officials believe that this would not be accepted by many communities as it may be interpreted as an effort to convert children to Islam. This made finding space for pullout classes more difficult and expensive, adding a need for additional tables and chairs.

The aim of this study was to replicate key components of previous studies, which were believed to be critical to effectively targeting instruction for weaker students, while also fitting the model to the Ghanaian context in a way that could be potentially scaled and sustainable in the long term. As described, earlier programs that rigorously evaluated targeting instruction at the child’s level were run by Non-Governmental Organizations, and as a result likely had more control over the implementation quality than a government would (Banerjee, Cole, Duflo, and Linden, 2005 in India and Dupas, Duflo, and Kremer, 2008 in Kenya). On the other hand, the TCAI Pilot described here was intentionally implemented through a relatively weak government education system using the existing monitoring systems. The TCAI pilot and study were designed to be nationwide with participating schools spread across the ten regions of the country - so as to encourage political acceptance at the study’s end. Thus, by design, any impacts of the TCAI program were measured in the context of a nationwide program using government systems, arguably indicating the potency of targeted instruction if it was to be mainstreamed in the government system.

\textbf{2.3 Intervention Design}

Based off of hypotheses formulated through existing evidence outlined above and education stakeholder discussions in Ghana, and further refined through a 12-week pilot in 2010, four program variations were designed to test the efficacy of targeting instruction to children’s learning levels, when run through the Ghanaian education system. While education stakeholders were interested in testing targeted instruction in Ghana, they also wanted to test the role of other potentially important

\textsuperscript{3} A more recent study, Sandefur et al, \textit{Scaling up what works: Experimental evidence on External Validity in Kenya Education}, found effects to disappear when a government implements the same program as a successful nongovernmental organization, demonstrating that the institutional context and the agency in charge of implementation can matter as much as the program design.

\textsuperscript{4} Renamed the Ghana Youth Employment and Entrepreneurship Development Agency (GYEEDA) in 2012, NYEP offered unemployed youth (18 to 35 year old), who are most often secondary school graduates, two year public service positions and a small ($80-100) monthly stipend. Among others, they are used by the Ghana Education Service to take on un-filled teacher positions (often in remote areas).
constraints, like Pupil Teacher Ratios (PTRs) and the lack of teacher training in instruction of basic skills. They wanted to know if teachers alone could be trained to target instruction at a student’s ability level, which would save costs and more likely to be accepted by the teachers union. They also had operational questions, like the best time to provide pull-out remedial education. Would conducting remedial classes after-school provide additional instructional time to children who need it the most? Or would students be less likely to attend as the program may not be regarded as integral to the school day and students may be eager to get home. Furthermore, an afterschool program would not provide teachers with the opportunity to have smaller classes focused on upper level learners. On the other hand, afterschool made sense because there was often limited space for during-school pullout classes and a temptation to use TCA time for other things.

The final program design included 4 interventions, involving the provision targeted instruction through teacher community assistants (TCAs) during or after school, professional teacher training on targeted instruction methods and reduced classroom size on learner outcomes. The program was implemented by the Ghana Education Services (GES), with the support of a small technical support team housed at Innovations for Poverty Action, but reporting to the Director of Basic Education and Steering Committee that comprised of members of each partnering institution. The TCAs were recruited through the National Youth Employment Program (NYEP), and the Ghana National Association of Teachers (GNAT) was a partner in the initiative and part of the Steering Committee. The piloted was implemented in grades 1, 2 and 3 in 500 primary schools, which were randomly selected in 42 Districts, to be nationally representative and divided into 4 treatment groups and a control group.

Below is a brief description of each treatment and related research question for each intervention, also summarized in Table 1:

**Intervention 1:** Provides in-school remedial classes through pull-out classes held by TCAs for students identified as having lower learning levels, with targeted instruction in basic math and reading, in English and the local language. TCAs provide these students with additional classroom attention, providing instruction on basic skills. This intervention allows for the measurement of targeted instruction for low level learners through pullout classes run by an assistant, while also providing the teacher with an opportunity to focus on upper level learners.

**Intervention 2:** Provides after-school remedial classes to lowest-level learners using TCAs. This intervention, while similar to intervention 1, provides an opportunity to assess the effectiveness of the remedial education program when implemented after school hours, providing students with an extra hour of instruction during a time when students are likely to have more focused time with TCAs.

**Intervention 3:** Provides an assistant to the school. The assistant conducts homework help to a half of the classes, which is randomly selected each day. The basic question here assesses the impact of reducing class size on quality, through the addition of a teaching community assistant.

**Intervention 4:** Provides training to professional teachers to develop skills in providing small-group instruction targeted at students’ actual learning levels. This intervention tests whether training on
targeted instruction can help teachers improve learning levels without the provision of a teaching assistant, and if so, what level students benefit from this training. 6

**Control** schools served as the comparison in measuring the treatment effects. No treatment was provided.

**Table 1: TCAI Interventions Components**

<table>
<thead>
<tr>
<th>Components</th>
<th>Targeted Instruction</th>
<th>Extra materials</th>
<th>Remedial pedagogy materials &amp; Extra Assistant</th>
<th>Smaller class size</th>
<th>Extra school hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Remedial classes for lowest-level students by TCA during school hours</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>2. Remedial classes for lowest-level students by TCA after school hours</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Smaller classes through the support of an extra assistant</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4. Training for professional teachers on targeted instruction</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The main inputs of the interventions described included: Teacher Community Assistants (TCAs) hired with sufficient education-levels and living in the school community; assistants and teachers trained; given a set of remedial materials 7 and supervised. Monitoring of the program was to be done through the government system by Circuit Supervisors, with the support of a small TCAI technical unit. 8

As a result, the following outputs were expected: Assistants would be in their assigned schools and teaching; targeted instruction (i.e. students tested and divided based on learning levels) would occur; teaching methods would focus on basic literacy and numeracy skills.

Each intervention was designed with the dual intent of testing the effectiveness of targeting the instruction of basic skills to the lowest-level learners as well as responding to stakeholders questions, which was crucial if the program was successful and to be scaled. Based off of a theory of change informed by stakeholders and earlier research, it was predicted that treatments with targeted

6 As described later, in year 1, teachers received training and resources alone. In year 2, grade 1 – 3 teachers were asked to reorganize their classes by ability levels, rather than grade, so as to better enable teachers to focus targeted instruction on basic skills.

7 Teacher learning materials included basic resources, which can be used to teach basic literacy and numeracy skills. Designed for the program based off of Pratham and local experience were delivered in boxes, included number, alphabet, word, sentence, paragraph and story cards, picture cards, manipulative items and slates. Teachers and TCAs were taught to use these items and make their own.

8 On average, there was 1 regional TCAI technical coordinator to support 100 schools.
instruction (Intervention 1, 2, 4) would be the most effective, but that Interventions 1 and 2, which implement targeted instruction through TCAs, would be the most effective as pull-out classes should make it easier for both assistants and teachers to focus their instruction on children’s levels. Nevertheless, it was also expected that implementation challenges would likely make targeted instruction less effective than if it was implemented by a more nimble non-governmental actor, like Pratham. For example, while previous experience indicated that the afterschool TCA treatment would allow for the most focused instruction time and likely higher outcomes, at the start of the study it remained an outstanding question as to whether this variation could be implemented at all through the government education system. Finally, it was expected that improvements from reducing class size through the addition of an assistant alone could take place (Duflo et al 2011 shows that adding a contract teacher, in itself, is effective, however those contract teachers were more qualified than the Ghanaians NYEP candidates), but that they would be less important (based on the evidence that reducing class size even for the entire day, in itself, did not results in improvements in learning levels).

3. Evaluation Design

The nationwide evaluation was designed to measure the impact of the TCAI Pilot after year 1 and year 2 of the program, and to collect data on inputs and outputs throughout the study. Key evaluation details, including the sampling, randomization and survey strategy, are detailed below.

3.1 Sampling

Sampling was done in two stages. First, 42 districts were randomly selected from 168 districts, with stratification by the number of districts per region, the proportion of rural population in a district, and the district deprived rank. Then, within each of the 42 districts, 11 to 12 schools were randomly selected, with stratified by PTRs and the urban/rural status of the school using the Education Management Information Systems (EMIS) data). Overall, urban schools were oversampled so that the differential impact of the program on urban and rural schools could be determined. When possible, an equal number of urban and rural schools were selected, and when there were not enough urban/rural schools, all urban/rural schools were selected. As a result, 3 districts did not have enough rural schools, and therefore have more urban schools in the sample. 21 districts did not have enough urban schools so these districts have more rural schools than urban schools represented in the sample.

At the school level, a maximum of 25 pupils from each grade 1 – 3 were randomly selected using the class registers’ photos collected during the preliminary baseline school survey. When grades had less than 25 pupils, all pupils were interviewed. In two schools, where class registers were not available and the class enrolment above 25, a list of all pupils in the class with serial numbers was created at the time

---

9 There are 170 Districts in the Country, and 2 were excluded from the sampling frame at the request of the GES, which was having issues with those Districts regarding NALAP.

10 GES has classified districts into deprived and non-deprived based on factors including access to educational facilities, performance, pupil teacher ratio, enrolment rates, etc. Then, Districts are given a deprived rank, on the basis of which a certain number of Districts get classified as “Deprived”. We used the deprived rank to stratify.

11 We randomly selected in which District we would select 11 or we would select 12 schools.
of the survey with the help of the Head Teachers, and a random table was used to do the random selection. When possible, an equal number of male and female serial numbers were selected. If pupils were found by the surveyors to have stopped attending the school during the sample selection process, they were replaced in the sample (if the class had more than 25 pupils).

3.2 Randomization

After the schools for the sample were selected, a baseline survey was conducted to gather essential data on the demographics and features of each of the 500 schools. The treatments were randomized across these schools, and balance checks were conducted to ensure that there were no statistically significant patterns in any of the 4 treatment groups or the control. Factors, such as average test scores (see Table 2), average pupil-teacher ratio, whether the pupil’s father can read (a measure of family literacy) and whether the pupil sleeps on a mattress (used an indicator for the pupil’s family wealth), were used to determined that the sample was statistically balanced across all four interventions and the control. There were no significant demographic or school-quality differences between any of the interventions. Thus, any differences in educational outcomes demonstrated between the interventions result from the programs, and not from any inherent differences or biases in the sample.

| alltests_ave | Coef. | Robust Std. Err. | t  | P>|t| | [95% Conf. Interval] |
|--------------|-------|------------------|----|------|-------------------------|
| treat1       | -0.3827214 | 1.299639 | -0.29 | 0.769 | -2.93616 | 2.170717 |
| treat2       | -0.3352623 | 1.423762 | -0.24 | 0.814 | -3.132569 | 2.462044 |
| treat3       | -0.1948054 | 1.375969 | -0.14 | 0.887 | -2.898212 | 2.508601 |
| treat4       | -1.714235  | 1.303602 | -1.32 | 0.189 | -4.275481 | 0.849711 |
| _cons        | 31.335453  | 1.004602 | 31.21 | 0.000 | 29.38076 | 33.3283 |

3.3 Measurement Strategy

The evaluation design includes a baseline at the outset of the program and two “endline” surveys to test pupil learning levels, in addition to three rounds of midline surveys, which gather data on the ground-level realities of the program, including attendance of pupils, teachers, and TCAs; classroom activities and teaching behaviour; and other features of the program uptake. These surveys are described in depth below:

---

12 A random table is a table where random numbers are provided in each of the column. For each school, surveyors are instructed to use one of the columns. The random numbers in this column are the serial numbers in the class register that should be selected.
3.3.1 Measuring final outcomes – learning levels

The baseline and the endline survey included similar sets of testing tools. Testing tools were developed in collaboration with the Assessment Services Unit of the Curriculum Research and Development Division of GES, and with the support of a psychometrician, Jeff Davis. They included:

- **Individually-administered oral tests** in English, math, and the local language (determined by the official language chosen by the National Literacy Acceleration Programme). These tests cover the critical objectives of the syllabus for grades 1 – 3, and one common test is used to test pupils across all grades. The baseline measurement tools were piloted on 300 children to test for validity and reliability (see Appendix Table 1). By request of the GES, these tools are aligned with the curriculum, and cover a range of skills, beyond what the targeted instruction training covered.

- **Written tests for Grade 3-4 students** (endline surveys only). These written tests were at grade level, although they included a section with simpler and lower grade items.

3.3.2 Measuring intermediary outcomes

Intermediary variables included whether the program was implemented as planned (e.g. TCA posted and present in school, teachers’ teaching practices) and variables which may have been impacted by the program’s intervention and contribute to improvements in learning outcomes (teachers’ and pupils’ attendance). These variables were measured through midline Survey: TCAI used surprise visits by surveyors to document the daily school activities at the schools in the sample. The first midline tools included the following instruments:

- **Survey questionnaires** for Head Teachers, Intervention 4 classroom teachers, and TCAs gathering data on the training and sensitization the school had received on TCAI, and on how the program is being implemented.

- **Observational surveys** on teacher and TCA attendance, teaching methods and activities in with schools in the sample.

- **Attendance and Remedial Attendance**: Data is collected on the school attendance of all selected pupils in the sample, and on their attendance in remedial classes for Interventions 1 and 2.

3.3.3: Baseline characteristics

The baseline survey also included survey questionnaires of the Head Teacher, classroom teachers, school management committees and parent teacher associations, community members, and pupils, to collect data on school infrastructure and quality, enrolment, teacher qualifications and background, and pupils’ socio-economic characteristics.

3.3.4: Data Collection Timeline and Strategy

The baseline survey was conducted in October-December 2010, with Grade 1-3 students. The first endline was implemented in November and December 2011 and included all pupils tested in grade 1 – 3 during the first year of the evaluation, who had progressed to grades 2 - 4. New pupils were added to the sample from the new P1 class, following the same procedure as the Baseline sampling method. The second endline was conducted in June and July 2012, and a new generation of students was again added. Grade 5 students (who were in grade 1 at the time of the baseline) were not tested, because
they had only benefite been in the program for three months, but Grade 4 students (those who were in Grade 2 at baseline) were, allowing to study the intervention’s effects one year after leaving the program.

Given high pupils’ absenteeism (about 30% on any given day), and given that it is frequent for children to change schools, one challenge facing the study results is attrition, and in particular differential attrition. The data collection strategy included testing children in school, but also tracking those children who were not found in school on that day, to minimize the effects of absenteeism. At endline 2, the survey teams were meeting only about 60 percent of the students present in school and hence had to track the remaining pupils in their homes or new residence if they had migrated. The team was able to finish the survey reaching 83 percent of the sample after an extensive two week post survey tracking exercise.

As shown in Annex, Table 2, attrition analysis indicates that it was mostly not differential, although it was 2 percentage points higher and significant for the Assistants Only group (Treatment 3).

3.3.5: Empirical Strategy

The empirical strategy is simple, since this is a randomized study. Using test scores as outcomes of interest, we compare each treatment group to the control group, and control for baseline results (for grade 3 and 4 as no baseline survey was conducted for grade 1-2 students), and the variables that were used in the stratification. For robustness checks, we also look at other regressions, with controlling only for the baseline survey, and baseline survey and Pupil Teacher Ratio, and find fairly similar results.

4. Implementation

Overall, the program design was followed: most assistants were recruited from the school communities, assistants and teachers were trained, materials distributed and students’ levels tested. However, all four interventions experienced important implementation issues, not surprisingly since they were implemented through the government system and at scale, with only a small implementation support unit.

4.1 Inputs

4.1.1 Assistants’ recruitment and characteristics

Overall, recruitment of TCAs went as planned. They were supposed to have been identified with the help of the School Management Committees or Parent Teacher Associations, and then interviewed by a panel comprising of GES and NYEP Representatives. The interview process was important to ensure that the TCAs fit the recruitment criteria and are not recruited on a political basis (a concern that was expressed by the GES). The majority of TCAs (80%) were interviewed before being hired, although these percentages differed by region and were higher in urban and non-Deprived Districts.

As a result, assistants largely fitted the recruitment criteria. They were supposed to be between 18 and 35 and to have completed secondary school with passing grades in the major subject areas (Maths, English and Science). They were also expected to speak, read and write the NALAP language assigned to
the school, and to live in the school community. According to the TCA interviews conducted at the first midline survey, most TCAs were indeed young, almost 90% of them lived in the school community, 66% were born in it, and 41% even attended the schools to which they were assigned. As this would imply, almost all TCAs spoke the local language assigned to their school. Educational attainment was fairly uniform across the cohort of TCAs—approximately 97% of TCAs had completed Senior Secondary School. However, all TCAs did not pass their graduation exams. TCAs were most successful in English and Science with a 89% and 85% pass rates; but, nearly a quarter of TCAs failed their Math exams.

Importantly, the percentages of candidates with passing grades dropped significantly for those TCAs who were not interviewed. This may simply indicate that in places with fewer options for TCA candidates, making interviews less relevant, lower quality candidates had to be recruited. However, it may also show the consequences of less rigorous recruitment in some locations.

4.1.2 Training and materials

In all three interventions involving TCAs (intervention 1, 2 and 3), 99 percent of TCAs attended the initial training in May 2011. In the teacher-led targeted instruction intervention (Intervention 4), 85 percent of teachers attended the training. Teachers and TCAs were nearly universally positive in their feedback on the trainings, rating the trainings “Good” or “Excellent” over 95% of the time in all categories: “Overall”, “Trainers”, “Lecture Content”, and “Group Work”. Interestingly, however, TCAs were more enthusiastic in their responses, responding “Excellent” significantly more often than Intervention 4 Teachers in all categories.

TCAs and teachers received materials, however the distribution was delayed. At the time of the first midline survey, that took place during the first term of the program, the materials had not been fully distributed, and only 22% of TCAs were found to have TCAI materials in their class at the time of visit. Even when they had received the materials, they were not using all of them, so in the second phase of the program simpler materials were designed and a smaller number of them.

4.1.3 Supervision

A monitoring system was initially designed to complement the existing mechanism. It involved distributing cell phones to Circuit Supervisors, head teachers as well as representatives of School Management Committees, and having them send text messages every day to report the attendance status of the TCAs. Representatives of NYEP as well as GNAT were also supposed to visit TCAs regularly. However, the proposed monitoring system did not work as anticipated. Circuit Supervisors (district level staff of GES), NYEP & GNAT rarely went to the schools. In the first few months of the program, School Management Committees checked-in in only 20 percent of the schools, and on average went only during the first 2 weeks. Head teachers were most likely to monitor TCAs, but the initial 70 percent response rate has dropped to about 30 percent. Over time, the program stopped relying on this monitoring system.

This means that for the most part, the program relied on the existing monitoring system (i.e Circuit Supervisors visiting schools once in a while). In addition, the IPA Technical Assistance Unit visited some schools, however there were only 4 Regional Coordinators for 400 program schools, so this was limited.
4.2 Outputs

4.2.1 Assistants’ attendance and time on task

In average and across the three interventions involving TCAs, over the program period TCAs were present 59% of the time (which is less than teachers, who are present in average 70% of the time). Attendance rates went down over time (see Appendix, Graph 1): From 72% at the first midline survey (the first trimester of the program), it fell to 53% at the third midline (the third trimester of the program). Overall, attendance was highest (67%) for Intervention 3 (TCAs but not targeted instruction). It was 61% for Intervention 1 (Targeted instruction during school hours), and only 49% for Intervention 2 (Targeted Instruction after school hours).

Time on task (defined here as the percentage of the time TCAs were found teaching a pull-out class) was even lower: overall and across treatments, TCAs were teaching a pull-out class 30% of the time (for comparison, teacher time on task, defined as percentage of the time teachers were found teaching the class, was found to be 34%). Time on task was highest for treatment 2 (36%), even though attendance was lower, probably because TCAs during schools were more likely to do other things (main reasons for not teaching the remedial class, apart from absence, included extramural activities, or teaching the whole class – about 10% of the times they were not teaching the remedial group). Time on task was lowest for treatment 3 (24%), however when they were not teaching a pull-out group, they were more likely to be teaching the full class. So overall, the assistants in treatment 3 were not teaching less, but they were more likely to teach the whole class.

Table 3: TCA Attendance and Time on Task

<table>
<thead>
<tr>
<th></th>
<th>Attendance (%)</th>
<th>Time on Task (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>59</td>
<td>30</td>
</tr>
<tr>
<td>During School</td>
<td>61</td>
<td>29</td>
</tr>
<tr>
<td>After School</td>
<td>49</td>
<td>36</td>
</tr>
<tr>
<td>Assistants-Only</td>
<td>67</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 4: TCA Time Attendance and Time on Task Over Time

<table>
<thead>
<tr>
<th></th>
<th>TCA Present (%)</th>
<th>TCA Time on Task (%)</th>
<th>Reason TCA class not happening: TCA Teaching Main Class (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treat 1</td>
<td>Treat 2</td>
<td>Treat 3</td>
</tr>
<tr>
<td>Midline 1</td>
<td>77.1</td>
<td>66.7</td>
<td>74.2</td>
</tr>
<tr>
<td>Midline 2</td>
<td>72</td>
<td>53</td>
<td>75</td>
</tr>
<tr>
<td>Midline 3</td>
<td>48.1</td>
<td>44.1</td>
<td>63.1</td>
</tr>
<tr>
<td>Midline 4</td>
<td>55.3</td>
<td>42.9</td>
<td>72.1</td>
</tr>
<tr>
<td>Midline 5</td>
<td>70.8</td>
<td>56.8</td>
<td>47.4</td>
</tr>
<tr>
<td>Midline 6</td>
<td>48.3</td>
<td>36.7</td>
<td>61.4</td>
</tr>
</tbody>
</table>
One of the reasons why assistants’ attendance decreased is the fact that their payments by the NYEP were very inconsistent – they were not paid for 8 months in a row. However, it does not explain everything, as assistants’ attendance varied a lot across Districts (while payments issues did not). Looking more closely, we also find that TCA attendance matches very closely with teacher attendance rates at the District level (see graph 2 – on average, teachers missed approximately 1.5 days of school per week, but this also varies a lot across Regions and Districts). This indicates that attendance and time-on-task are likely related to a broader supervisory issue at district-level, rather than only a TCA programmatic concern. Indeed, as mentioned above the interventions, relied on the existing supervision mechanisms, yet administrative data and midline surveys collected by TCAI show that supervision by circuit supervisors was low. Consistent with Sandefur et al (2013), “implementation issues and political economy forces put in motion as the program went to scale”.

4.2.2 Targeted instruction

Overall, the TCAs and teachers seemed to have adhered to the TCAI basic methodology: During the first midline, 99 percent of TCAs in Interventions 1 and 2 reported testing students to form remedial groups, while 90 percent of TCAs for Intervention 3 reported allocating students into sections randomly. This indicates that the TCAs correctly followed the basic premise of the treatments, with Interventions 1-2 focusing on the weakest pupils and Intervention 3 TCAs choosing pupils at random to divide the class in half.

As a result we expected baseline test results to be closely correlated to whether a child was in fact assigned to a remedial class. Indeed, at the first midline, over 52% of pupils assigned to the remedial class scored below the median test results. However, while statistically the prediction is very significant, it is not as high as one would have expected. One reason is probably that overall levels are very low, so more than half the children are below grade level, and the baseline test was more detailed than the simple test used by the TCAs to assign pupils in a remedial class.

According to information collected during midline surveys, in averagely about 35 – 40 percent of students were assigned to remedial classes per trimester, in During and After School interventions. This assignment was dynamic and took place every trimester and possibly even more often. As a result, throughout the program between 73 and 86% of students were assigned at least once to the remedial class. This percentage was a bit higher for the After School class, however students were also more likely to be missing the after school class.

In the teacher-led intervention, teachers were not consistently implementing the targeted instruction approach. According to teachers, 80% divided their class into ability groups, and of these teachers, 95% tested pupils in order to form the appropriate groups. However, only 47% of teachers were able to show the test results to the surveyors, and according to midline observation surveys, they split their classes by ability levels only 15% of the time (the 85% of days where it did not happen also include days where the teachers were absent, 30% of days in average). Overall teacher time on task (time spent teaching) was 34%, so teachers implemented the targeted instruction approach a bit less than half their teaching time.
4.2.3 Teaching behaviors

The quality of teaching by assistants was low at first, and the implementation support team and the Steering Committee realized that they would need more training than initially planned for. Those issues were partly addressed through a revised, District-level training plan after the first endline survey.

During the midline surveys, surveyors were asked to rate the teaching quality, by assessing to what extent the TCAs or teachers were following the methodology that they had been taught to teach new words, operations and the place value concept.

Overall, TCAs rated fairly well. According to surveyors, 61% of them were using the right methodology to teach words and place value; 73% to teach operations; but only 48% used the phonetic method appropriately. Data on TCA teaching quality showed more variations across regions than with interventions\(^\text{13}\). However the two remedial interventions were rated higher than the random split TCAs, which was expected since only the remedial TCAs received training focused on teaching the weakest.

According to our midline observations, teachers who received the Targeted Instruction training also changed their teaching behaviors. Compared to teachers in the control group, teachers in treatment 4 were more likely to use the new reading and place-value teaching methods. For example, 38% of teachers in treatment 4 used the phonetic method, compared to 17% in the control group, and 52% of teachers in treatment 4 used the place-value teaching method, compared to 29% in the control group.

5. Results

5.1 Teachers and students’ attendance and efforts

We observed no difference in students’ absenteeism (32% in average) between all five groups.

There is no statistically significant difference in attendance levels between the three interventions involving assistants, and the control. This is important, because one fear about this program was the possibility that teachers would start coming less and ask their assistant to teach the class for them. Interestingly, teacher Time on Task even increases in a significant manner, only for the Assistant only group, by 8%.

In Intervention 4 (teacher-led targeted instruction), across the 6 midline surveys teacher attendance was significantly higher (77% attendance rate) compared to the control group (71% attendance rate), and the percentage of time spent teaching while in school was 11 percentage points higher (from 62% to 73%).

\(^{13}\) The TCA quality rating was given by the person observing the class based on his/her of the TCAs’s teaching delivery.
We believe that we can rule out that the attendance effect was driven by additional school visits by the TCAI technical assistance team or the evaluation team surveyors, because the same level of monitoring occurred in Intervention 1 and 3. However, it could be that since the school visits were more focused on the teachers in the Teacher-led Intervention, attendance rates were more likely to increase in this group than in others. However, the midline surveys were not different in any of the groups, the technical assistance team consisted only of 1 Regional Coordinator for 100 schools so their visits were not frequent, and there was no consequence for absenteeism, so it is likely that the increased attendance rates were due to the intervention rather than those visits.

5.2 Test scores: Overall Impact

5.2.1 Overall effects across Grades 1-4

All results outlined below are Intention to Treat Estimates.

The endline survey tested students in grades 1 to 4. When taking all grades into account, the assistants-led targeted instruction intervention resulted into a 0.08 standard deviation increase when implemented during school hours (significant at the 5% level), and 0.075 one when implemented after school hours (significant at the 10% level), while the other two interventions did not have a statistically significant effect.

Since grade 1 students had been in the program for just a year, and grade 4 students had been out of the program for 1 year (after benefiting from the program for 1 year and 1 term), it is relevant to look separately at grade 2 and 3 students who had been in the program for 2 years or more. Indeed, effects on grade 1 pupils are small and insignificant.

5.2.2 Overall effects for students in program for 2 years (grades 2-3)

For grade 2-3 students, the assistants-led targeted instruction intervention resulted into a 0.09 standard deviation increase when implemented during school hours (significant at the 5% level), and 0.08 one
when implemented after school hours (significant at the 10% level), while the other two interventions did not have a statistically significant effect.

Most of this effect on the overall test comes from the grade 3 students. Effects for grade 2 students are only visible for the most basic skills (e.g. alphabets). Test scores in grade 3 are 0.16 standard deviations higher than in the control group for the During School group, and 0.14 standard deviations higher than in the control group for the After School group, both significant at the 5% level. Test scores are also higher by 0.1 standard deviations in the Teacher-Led group, but only at the 10% significance level.

Grade 3 pupils also performed a written test, on which impacts are positive and significant for the two Assistants-led Targeted Instruction interventions (0.13 standard deviation during school, 0.18 after school – see Graph 1). The Assistant Only group also had a 0.12 standard deviation impact, but only at the 10% level. The Teacher-led intervention did not lead to improvements.

Table 7: Overall Oral Test Results

<table>
<thead>
<tr>
<th></th>
<th>Grades 1-4 Oral Test</th>
<th>Grades 2-3 – Oral Test</th>
<th>Grade 3 – Oral Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>During School</td>
<td>0.083*</td>
<td>0.092*</td>
<td>0.160**</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.047)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>After School</td>
<td>0.075+</td>
<td>0.078+</td>
<td>0.144**</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.045)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Assistants Only</td>
<td>0.047</td>
<td>0.032</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.042)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Teacher-Led</td>
<td>0.036</td>
<td>0.043</td>
<td>0.097+</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.046)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

5.2.3 Persisting effects for Grade 4 students

The effects on grade 4 pupils, who have not been in the program for a year and overall have participated in the intervention for one year and one term, is also positive. For the During and After School interventions, test scores are respectively 0.128 and 0.11 standard deviations higher than in the control group – which is only 0.03 standard deviations lower than the effect on grade 3 students. This means that the intervention has persisting effects. The effects also persist on the written test, but only for the English easier section (section 1), for which test scores were 0.14 and 0.11 higher than the control group for the During and After school groups. This section of the test included items of grade 2 level, as opposed to other sections which were closer to grade 4 level. Given that the grades 4 pupils had exited the program about a year ago, it indicates that maybe only the more basic skills in reading gained persist.

Surprisingly, effects for the Assistants Only group were higher than for any other grades: 0.165 standard deviations. It might be that in the Assistants Only group, assistants were more embedded into the school and used across grades even though they were supposed to focus on grade 1-3. If that is the case Grade 4 students may have continued to benefit from their presence. However we have not verified this hypothesis, although anecdotal evidence suggests this might be the case.
5.3 By Subjects and Targeted Sections

Overall, results are higher for literacy than for math, and higher for English than for Local language. For all grades, scores in During and After School groups are 0.08 and 0.09 standard deviations higher than the control for Literacy, but not statistically significant for Math. This is consistent with a strong focus on literacy by the Steering Committee and the technical assistance team after the results of the first endline survey, which had revealed better results in Math.

In grades 3 and 4, there are statistically significant effects for both Math and Literacy for all three Assistants interventions, but effects are higher for Literacy. Effects are also significant in Literacy for the teacher-led Intervention. English impacts are higher than Local Language for the After School intervention, while Local language impacts are higher During school. It is likely that because of the government’s focus on local language, during school the TCAs were instructed by teachers to focus more on the local language, where as they had more freedom after school.

While the tests covered all components of the curriculum, at the request of the Ghana Education Service, the TCAs were trained to teach a certain set of skills: basic reading and computation skills. The English test included 3 sections: 1) Sections on Listening Reciting & Conversation; 2) Grammar, and Vocabulary and 3) Reading and pre-reading. The Math test sections included: 1) Numbers and Fractions; 2) Computations and Operations, and 3) Geometry, Measurement and Data. Given the focus of the targeted instruction training, we expected Literacy Section 3 to be mostly impacted, and Math Section 1 and 2. Results confirm this hypothesis: When looking at results for all grades combined, they are only significant for the sections 3 of the English and Literacy tests, and Section 2 of the Math test. Section 1 of the Math test does not yield statistically significant results. We hypothesize that this is because math baseline results are better for math than for literacy, so students were more likely to already score fairly well at number reading. Among these targeted sections and across all 4 grades, results range from 0.7 standard deviations (Math S2, After School group) to 0.11 standard deviations (English S3, After School group).

We also expect that across sections, results should be higher for the easiest questions, or the most basic skills. This is indeed what we find when looking at grades 2 and 3 together, or grade 2 alone, and this is true for all three targeted instruction interventions.

Table 8: Tests Results by Section and Very Basic Skills (Grades 2-3)

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Local Language</th>
<th>Math</th>
<th>Math S2</th>
<th>Eng S3</th>
<th>Loc S3</th>
<th>Lang S3</th>
<th>Very basic Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>During School</strong></td>
<td>0.090+</td>
<td>0.066</td>
<td>0.066</td>
<td>0.084*</td>
<td>0.076</td>
<td>0.119*</td>
<td>0.104*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.054)</td>
<td>(0.043)</td>
<td>(0.040)</td>
<td>(0.047)</td>
<td>(0.052)</td>
<td>(0.049)</td>
<td></td>
</tr>
<tr>
<td><strong>After School</strong></td>
<td>0.111*</td>
<td>0.034</td>
<td>0.037</td>
<td>0.073+</td>
<td>0.126**</td>
<td>0.088+</td>
<td>0.120**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.055)</td>
<td>(0.043)</td>
<td>(0.038)</td>
<td>(0.044)</td>
<td>(0.052)</td>
<td>(0.045)</td>
<td></td>
</tr>
<tr>
<td><strong>Assistants Only</strong></td>
<td>0.042</td>
<td>0.027</td>
<td>0.018</td>
<td>0.026</td>
<td>0.045</td>
<td>0.028</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.051)</td>
<td>(0.041)</td>
<td>(0.037)</td>
<td>(0.045)</td>
<td>(0.048)</td>
<td>(0.045)</td>
<td></td>
</tr>
<tr>
<td><strong>Teacher-Led</strong></td>
<td>0.049</td>
<td>0.027</td>
<td>0.021</td>
<td>0.051</td>
<td>0.101*</td>
<td>0.107*</td>
<td>0.129**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.054)</td>
<td>(0.043)</td>
<td>(0.039)</td>
<td>(0.044)</td>
<td>(0.050)</td>
<td>(0.045)</td>
<td></td>
</tr>
</tbody>
</table>
In grades 3-4 only, results are also highest in Math Section 2 and Literacy Section 3. In these sections, impacts range from 0.12 standard deviations (After School group, Math S2) to 0.18 standard deviations (local Language S3, During School hour). However, for grades 3-4 there are also statistically significant, positive effects for English Sections 1 and 2, for the During and After school groups (from 0.12 to 0.14 standard deviations). Effects on the most basic skills are also not higher than the effects on the overall test. This is consistent with the fact that there were effects on the written test, and confirms that in higher grades, the intervention did not only affect the most basic skills but also resulted in broader effects.

Table 9: Tests Results by Section and Very Basic Skills (Grade 3)

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Local Language</th>
<th>Math</th>
<th>Math S2</th>
<th>Eng S3</th>
<th>Loc S3</th>
<th>Very basic Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>During School</td>
<td>0.150*</td>
<td>0.125*</td>
<td>0.119*</td>
<td>0.147**</td>
<td>0.126*</td>
<td>0.176**</td>
<td>0.118*</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.063)</td>
<td>(0.053)</td>
<td>(0.052)</td>
<td>(0.058)</td>
<td>(0.065)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>After School</td>
<td>0.167**</td>
<td>0.106+</td>
<td>0.087+</td>
<td>0.130*</td>
<td>0.194**</td>
<td>0.150*</td>
<td>0.164**</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.063)</td>
<td>(0.053)</td>
<td>(0.053)</td>
<td>(0.058)</td>
<td>(0.065)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Assistants Only</td>
<td>0.070</td>
<td>0.064</td>
<td>0.048</td>
<td>0.065</td>
<td>0.076</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.059)</td>
<td>(0.049)</td>
<td>(0.049)</td>
<td>(0.056)</td>
<td>(0.060)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>Teacher-Led</td>
<td>0.104+</td>
<td>0.082</td>
<td>0.054</td>
<td>0.097+</td>
<td>0.153**</td>
<td>0.161**</td>
<td>0.142**</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.062)</td>
<td>(0.051)</td>
<td>(0.051)</td>
<td>(0.055)</td>
<td>(0.060)</td>
<td>(0.050)</td>
</tr>
</tbody>
</table>

5.4 Class Size, Heterogeneity of Initial Learning Levels, and Initial Learning levels

Since the During School and the Assistants Only interventions involved splitting the class for some time each day, one could expect this to affect large classes more than others. However, we don’t find any differences.

The targeted instruction interventions should work best in the presence of a heterogeneous classroom. One way to test this hypothesis is to look separately at the effect of these interventions on multigrade classes. At the time of the study multi-grade classes (which typically bring together Grades 1-2, and Grades 3-4) represented about a fifth of the sample schools. These classes exist because the GES has a minimum Pupil to teacher and some schools don’t have enough students in each grade to fit this criteria. Anecdotal evidence however suggests that inadequate supply of teachers, particularly to deprived communities, are the major cause of multi-grade classes.

Consistent with the hypothesis that a targeted instruction program should work best in the most heterogeneous environments, we indeed find the highest effects among those schools, in particular for

---

14 Multi-grade classes refers to the presence of more than a grade sitting together in a classroom for the normal school teaching.
During School group. This is to be expected, since in this group students are separated based on
ability on a regular basis (where as they still remain together during school hours in the After School
group, and the intervention was not as consistent for the Teacher-Only group). Overall and across all
grades, the effect of the program for the During School group was 0.2 standard deviations higher (at the
10% level) for multigrade classes than in other classes. It was 0.1 standard deviation higher for the After
School group, although this difference is not statistically significant.

Interestingly, this difference between multigrade and other classes was even higher for grades 1-2. For
these grades, test scores in the During School group were 0.28 standard deviations higher than in the
control group, up to 0.43 standard deviation higher for the Local Language test.

Multigrade classes tend to be in more rural and possibly more deprived communities, which could have
driven these differences, but we also look at rural and urban schools, and deprived and non-deprived
districts separately, and do not see any differences. This, and the fact that these differences are most
important for the During School group, indicates that this is about the heterogeneity in the classroom,
which intervention addressed the most directly.

We don’t find any significant differences between children below and above baseline survey results. This
is not surprising, since as mentioned above, the correlation between being assigned to the remedial
class and being below the median was not high at the first midline survey, and up to 80% were assigned
to the remedial class at least once throughout the program.

6. Interpretations and Policy Implications

6.1 The role of implementation

This program was mostly implemented through the government system. The TCAI technical assistance
Unit worked with the GES on program design and provided some limited monitoring to schools,
however the Unit only consisted of 4 Regional Coordinators, so 1 for 100 schools. The interventions
therefore mostly relied on the existing monitoring system, which involves Circuit Supervisors visiting
schools on a regular basis. However, while they are supposed to visit schools 3-4 times a term, these
visits are in reality much less frequent. As a result the program was faced with important
implementation challenges, consistent with issues described in Sandefur (2013), which inevitably
affected the interventions’ impacts. While we did not randomize implementation variations, we can look
at differences across regions and changes over time give us a sense of the extent to which
implementation affected for the interventions’ impacts.

6.1.1 Regional Differences and Variation in implementation quality

When looking more closely at the evaluation results, we found a high variation in impacts across the 10
regions of Ghana, mostly for the assistants interventions. The variation in impacts between regions was
at times even higher than the difference between the average impacts of the four interventions. For
example, in Upper West, the impact of the During School group was 0.47 standard deviations for grade 3-4 students, whereas it was non-existent in Greater Accra.

Not surprisingly, the variation in impacts from region to region overall seems to follow variations in implementation quality (although we did not randomized implementation quality factors, so cannot make rigorous attributions). For example, the four regions with highest effect sizes for the remedial intervention during school and after school were also the ones with some of the highest TCA attendance. Generally Greater Accra and Upper East TCAs (2 regions with the lowest impact results) were rated among the poorest in terms of teaching quality. The regions that performed the best were the Upper West, Western and Central Regions. Apart from TCA attendance, other factors common in these regions included they having the highest rates of TCAs living in the school community (for intervention 1), and higher percent rates of teaching above 4 hours (self-reported). Also TCAs in these 3 regions reported the highest rates of in school support. All intervention 1 TCAs interviewed in the Upper West and Western region reported that their Teacher mentor visited their remedial class in the past week. The Central also intervention 1 TCAs also reported high rates of the Teacher Mentor visiting the TCA remedial class. Further the 3 regions were among the highest reported to having their circuit supervisors ever visiting the school since the TCA joined the school.

This highlights the importance of monitoring and mentoring on a regular basis for such a new program to be successful.

6.1.2 Implementation Quality over time

Impacts for the same cohorts of students increase over time, which means that the interventions continue to produce results. However, as mentioned above the quality of the assistants-led targeted instruction interventions decreased over time, as measured by their attendance and time on task.

Ideally, the best way to ascertain this would be to compare impacts students at same grade level, at Endline 1, and at Endline 2, after the same amount of exposure to the program. The timing of our surveys don’t allow to do this, however we can compare those students who were in the middle of P2 at Endline 1 (they had had started the program the last term of P1, and had been exposed to the program for 13 weeks at EL 1), and students who were finishing in P2 at EL 2 (they had had started the program at the beginning of grade 1, and had been exposed to the program for 38 weeks at EL2). If the program quality had remained the same, we should expect higher impacts for those P2 students who have been exposed to the program the longest. However, this is not the case: At endline 1, test scores for P2 students were 0.08 standard deviations higher than the control group (statistically significant at the 10% level), and at endline 2, for P2 students they were only 0.04 higher and not statistically significant, although this cohort had been part of the program for longer. This is probably due to the fact that time on task has significantly decreased since the beginning of the program, so that overall exposure was in fact not higher for P2 students at EL2.

As mentioned above, the positive results are mostly driven by grades 3-4 students. This could partly be explained by the fact that there is more variation in learning levels and that targeted instruction might therefore work better in higher grades. However, as mentioned there were positive results for these P3 students, when they were early P2 students. So these results seem mostly driven by the fact that the P3-
4 students at endline had benefited from the program when it was best implemented and when their exposure to it was the highest.

On the other hand, for treatment 3 and 4, effects in Endline 2 for P2 students are higher than they were for P2 students at Endline 1, indicating that these two interventions did not get worse, and possibly improved. It seems that assistants in the Assistants Only group took more time to figure out the right way to implement, that over time they became a more integral part of the school, and their attendance has decreased less. As a result, the intervention’s quality seems to have increased over time. For the teacher-led intervention, as mentioned above there was a conscious decision to change the way that targeted instruction was delivered.

6.1.3 Implications

There are a number of policy implications from understanding these implementation issues. First, not surprisingly the way a program is implemented can affect whether a successful proof of concept will have a positive effect or not. This means that identifying monitoring mechanisms that will ensure a certain level of exposure to the program is key. Yet, designing an innovative monitoring system is challenging, as described above, so this needs to be done in a way that can realistically be implemented through the system. This is an important area of future research.

Another important policy implication is that whether TCAs or teachers, a one-time training is not enough and refreshers, and if necessary course corrections in the design are necessary (as was done here with the Teacher-Led intervention).

Finally, the fact that the effects were higher and mostly concentrated on grades 3 and 4 who participated in the early days of the programs, indicates that more intense exposure (high exposure over a shorter period of time) is likely to yield higher impacts than limited exposure over time. Programs that take this, and the difficulty of setting up effective monitoring systems over time, such as intensive camps at the beginning of the school year to catch up students who lag behind, may be more realistic and effective.

6.2 Which interventions should be recommended in Ghana?

6.2.1 Overall cost-effectiveness

An important consideration for policy makers determining which interventions to roll out is of course costs. All four arms of the TCAI program increased student learning by varying amounts, and they also incurred different costs, meaning that some arms achieved learning gains more cost-effectively than others.

The bulk of the costs of TCAI directly fund the components that drive program effectiveness: training for TCAs who provide targeted instruction in basic skills to small groups of pupils. When program costs are projected over 10 years of government implementation, training in targeted teaching methods accounts for around 75 percent of the total cost. This includes initial training of the “master trainers,” training of teachers and TCAs in the targeted instruction methods, and refresher trainings to keep the curriculum fresh. Because TCAI takes advantage of community youth who would already be paid through GYEDA, salary costs are limited to program management, and account for less than 10 percent of total costs. The
remaining 15 percent of program costs cover teaching materials, administration, and program monitoring\textsuperscript{15}.

Even though the additional costs of the TCA program did not include salaries, the Teacher-led intervention was less expensive, as it required training fewer people (teachers were trained as well in the TCA interventions) and did not involve recruitment costs. The Teacher-led Intervention cost $12.6 per child, the During and After School interventions cost $20.2 per child, and the Assistants only one cost $19.6.

Although it is slightly cheaper to provide either small group instruction through normal curriculum TCAs, or targeted lessons through classroom teachers, these interventions do not improve test scores as much as the combination of targeted instruction for low performing students and community assistants. As a result, providing targeted instruction to the lowest-performing pupils through teacher community assistants (TCAs) is more cost-effective than either smaller-group instruction or targeted lessons alone. We estimated that per $100 spent, the During School intervention generates a total of 1 Standard Deviation Improvement, for only 0.72 Standard Deviations in the case of the Teacher-Led group, and 0.8 in the case of the Assistants Only group.

6.2.2. Adding an assistant: Does Targeted Instruction matter?

Adding an assistant alone, without targeted instruction, had some positive effects on learning levels, although these effects were concentrated on grade 4 students. Teachers Time on Task was also higher in this group.

Overall, however, effects are higher when assistants were specifically trained to identify and focus on the students who lag behind, confirming other studies that demonstrated that targeting instruction at the child’s level is critical to improve learning levels. Additional costs of providing this training are marginal, so the cost-effectiveness of this approach is also higher.

6.2.3 Teacher-led vs Assistant-led Targeted Instruction

Adding an assistant means adding an extra resource, and is likely to be more expensive and more politically challenging than changing teacher practices, which is why it was critical for the Ghana Education Service that the study includes an arm where targeted instruction is provided by teachers alone, without assistants. However, the teacher-led targeted instruction intervention led to smaller effects than adding an assistant to teach remedial classes. This is likely explained by the low compliance by teachers to the intervention design: despite low attendance of assistants, they were more likely to be present than the teachers to change their teaching practice. Hence, without any changes to the monitoring system or strong enough incentives for teachers to change their methods, then it is likely that providing assistants to deliver targeted instruction will be more effective.

\textsuperscript{15}These Cost Effectiveness estimates taken into account the costs and impacts of the program that was evaluated and are therefore only indicative, and most useful to compare the four interventions among each other. It could be that at scale, these costs change – the program may need additional monitoring costs to be most effective, but costs may also go down with additional economies of scale. Including the costs of the TCA salaries would also lead to higher costs for the TCA interventions.
However, the fact that the teacher-led intervention has positive effects despite such low compliance is noticeable. This indicates that if compliance was higher, impacts would also likely be higher. Given the complexity of adding additional resources to the school system, it is critical to investigate realistic and cost-effective ways to increase teachers’ willingness to introduce classroom changes which have positive effects. In a further study, we are investigating the effects of enhancing monitoring and mentoring of teachers by Circuit Supervisors.

6.2.4 Understanding the role of school environments: During or after school?

By definition, programs at scale are implemented through a variety of contexts, whether they are different regional contexts, or school environments. To be scalable and to be made a national approach or policy, the design of programs must be standard and simple enough – or the issues described earlier of large-scale implementation would be worse. On the other hand, for the impact of the program to be maximized, its design should ideally be adjustable to a variety of school contexts. In order to do this, it is critical to understand the role of school environments and how they affect the impact of the program.

The evaluation was designed with this goal in mind, involving a large enough sample size to be able to look at impacts on sub-groups, and being nationally representative. Looking at impacts of the different interventions on sub-groups provided very useful insights on how program delivery could be adapted to school contexts. For example, one of the initial questions and debates was whether remedial classes should be delivered during or after school hours. In average, the effects are fairly similar. However, this hides some important variations depending on contexts. Overall, it seems that in school environments that were more conducive for the remedial classes (e.g. schools with more extra space and extra furniture), the impact was higher for the during school intervention. When this was not the case, the class was less likely to take place because of logistical issues. Ultimately, the recommendation should be that the class be organized according to the school context, in a way that maximizes the likelihood of the remedial class to happen.

As mentioned above, the During school remedial intervention also had higher impacts than after school in schools with multi-grade classes. In schools with multi-grade classes, the effects of remedial classes during school hours was about 0.3 standard deviations, 0.15 standard deviation higher than the average effect and .1 standard deviation lower than for the after school treatment in multi-grade environments. This confirms the importance of teaching children at their ability levels and the relevance of this approach, since heterogeneity with learning levels must have been higher in multi-grade classes compared to single grade classes. It also demonstrates that when providing teaching assistance in this context targeted instruction may be particularly effective to improve learning levels and should be highly recommended.

Measuring variations in how the program affected students differently can also inform targeting. For example, girls were found to benefit more from the after school program than boys. We found that girls’ local language test scores increased by 0.12 standard deviations more than boys’ scores. Analyzing baseline data on after school activities that affected learning, girls were also found to be more involved in everyday activities like fetching water and firewood, 20 percent for girls compared to 15 percent for boys, while boys were more involved in seasonal activities like farming. Also girls report to have their mothers help them with their homework while boys were more likely to have their fathers’ help, who
are more likely to be literate. This indicates that the after-school program, especially in lower income areas, can provide an important support for female students.

Conclusion

The TCAI evaluation confirms that the targeted instruction at a student’s level is an effective approach to improving learning levels in literacy and numeracy in lower primary, and when implemented with some consistency and quality the positive effects can be quite high.

This evaluation also confirms that Sandefur et al (2013) rightly warned that potential impacts of the targeted instruction approach designed by a non-profit would be dampened by implementation issues intrinsic to government-led programs. And while we cannot measure the impact of these issues, the TCAI Pilot faced significant challenges related to administration, management and accountability, all of which are inherent to a weak government education system. Nevertheless, this study demonstrates that quality and consistency can be achieved locally in such a system, and provides data that illustrates key pre-conditions necessary to run effective programs focused on school quality.

It is our hope that this work will encourage, and help to inform, investment in system wide improvements that may not lead to immediate impacts. As well as more focused research on how to most efficiently and effectively strengthen government implementation systems, providing an enabling environment through for programs focused on quality.
References


Annexes

Table 1: Test reliability

<table>
<thead>
<tr>
<th></th>
<th>Cronbach's Alpha</th>
<th>Average Inter-item Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tests All Sections</td>
<td>0.93</td>
<td>0.18</td>
</tr>
<tr>
<td>English Test</td>
<td>0.88</td>
<td>0.27</td>
</tr>
<tr>
<td>Section 1</td>
<td>0.81</td>
<td>0.38</td>
</tr>
<tr>
<td>Section 2</td>
<td>0.67</td>
<td>0.25</td>
</tr>
<tr>
<td>Section 3</td>
<td>0.77</td>
<td>0.30</td>
</tr>
<tr>
<td>Local Language Test</td>
<td>0.82</td>
<td>0.18</td>
</tr>
<tr>
<td>Section 1</td>
<td>0.75</td>
<td>0.30</td>
</tr>
<tr>
<td>Section 2</td>
<td>0.60</td>
<td>0.20</td>
</tr>
<tr>
<td>Section 3</td>
<td>0.75</td>
<td>0.27</td>
</tr>
<tr>
<td>Math Test</td>
<td>0.84</td>
<td>0.20</td>
</tr>
<tr>
<td>Section 1</td>
<td>0.70</td>
<td>0.25</td>
</tr>
<tr>
<td>Section 2</td>
<td>0.78</td>
<td>0.31</td>
</tr>
<tr>
<td>Section 3</td>
<td>0.52</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 2: Attrition by Intervention

<table>
<thead>
<tr>
<th></th>
<th>attrition_cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) treat1</td>
<td>0.0184 (1.28)</td>
</tr>
<tr>
<td>treat2</td>
<td>0.00364 (0.26)</td>
</tr>
<tr>
<td>treat3</td>
<td>0.0297* (2.04)</td>
</tr>
<tr>
<td>treat4</td>
<td>-0.0173 (-1.06)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.804*** (76.05)</td>
</tr>
</tbody>
</table>

Observations: 22357

T statistics in parentheses

"* p<0.05    ** p<0.01    *** p<0.001"
Graph 1: TCA Attendance over time

Graph 2: Written Test Scores, P3

**Written Test Scores P3 (SD)**

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treat1</th>
<th>Treat2</th>
<th>Treat3</th>
<th>Treat4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>0.13*</td>
<td>0.18*</td>
<td>0.12+</td>
<td>0.07</td>
<td></td>
</tr>
</tbody>
</table>