Learning to Learn by Learning to Teach: Pre-Analysis Plan

Nava Ashraf Abhijit Banerjee Vesall Nourani

October 2, 2020

Abstract

This pre-analysis plan outlines our approach to studying the impacts of a novel teacher training program in Uganda. We partner with a Ugandan NGO that trains teachers by facilitating a yearlong practical learning and research experience, interspersing intensive periods of study during school holidays with in-service coaching. The teacher-trainers guide teachers using a curriculum titled *Preparation for Social Action* that, in part, equips teachers with scientific capabilities. The curriculum combines text that deepens a participant's understanding of the process and methods of science while simultaneously engaging with practical activities that facilitate analysis of natural and social processes in one's local community. We tailor the pre-analysis plan around studying the impacts of this training on (upper) primary schools in Jinja district, focusing on changes in 1) teacher pedagogy, motivation and effort, and 2) student learning. We hypothesize that the program enables teachers to implement an exploratory pedagogy that improves traditional student learning outcomes and also facilitates deeper learning in ways that makes knowledge directly applicable to increasing the well-being of the teachers' and their students' communities.

Contents

1 Introduction

Uganda received considerable international attention for being the first African country to adopt a policy of free universal secondary education in 2007, ten years after adopting free Universal Primary Education. While removing primary and secondary school fees should have increased participation in Post-Primary Education, in 2016, only 35.5% of students reached the last grade of primary education and only 59% of primary-school leavers effectively transitioned to lower secondary school.[1] Thus, efforts to improve the quality of and increase participation in secondary education must consider the entire schooling pipeline.

The Ugandan Ministry of Education and Sports (MoES) has placed increasing emphasis on increasing the number of qualified and capable teachers through in-service training. There is a growing sense that pedagogical approaches are overly hierarchical and disconnected from labor market outcomes among youth — unemployment and labor force participation rate among youth aged 18 to 30 were, respectively, measured at 13.3% and 57.3% in the 2016/2017 Uganda National Household Survey (Uganda Bureau of Statistics, 2018, 2019). Simultaneously, high rates of teacher absenteeism may reflect low teacher morale, a sentiment shared by many stakeholders in Uganda's education.

We explore whether a novel teacher training can sufficiently shift pedagogy and other dimensions of teacher quality in a manner that can overcome these challenges. Our partner organization in Uganda, Kimanya-Ngeyo Foundation for Science and Education (KN), utilizes a curriculum titled *Preparation for Social Action* (PSA) to engage *teachers*. The training is best conceived of as an experience in which participants deepen their understanding of the process and methods of science while simultaneously engaging with practical activities that facilitate analysis of natural and social processes in one's local community. More concretely a subset of the modules in this curriculum help teachers 1) describe the world with increasing clarity and precision 2) practice and develop the skill of making inference from observations, 3) formulate and express hypotheses and

[1] UNESCO – Institute for Statistics, find further information at `http://uis.unesco.org/en/country/ug`.

evaluate evidence with increasing clarity of language and thought, and 4) connect local knowledge and practice with the accumulated bodies of global scientific knowledge to advance development. A series of modules within the PSA curriculum are immediately relevant to processes of community life and focus on agriculture, education, health, environmental issues and even include exercises that facilitate reflection around the benefits and disadvantages of social norms to one's community.

The PSA curriculum combines content knowledge with practices and questions that guide discovery. We share four, among many more, examples of how the PSA curriculum does this using concepts in agricultural modules. First, in a unit titled "Planting Crops," teachers are asked to collect leaves from diverse plants in their environment, observe the leaves' structure, and classify them using tools of observation. They are asked to compare classification decisions with other teachers to ensure that their decisions are sound and reasonable to their peers. Second, participants engage in experiments aimed at comparing the degree of soil erosion using diverse soil-preservation techniques; e.g., comparing mulching to non-mulching practices — a clear experiment that demonstrates one means of making inferences. Third, teachers learn to measure soil pH levels, using locally available resources, of soils with different degrees of agricultural inputs — allowing them to describe the world with greater clarity and precision. Fourth, they observe leaf and root structures of plants to identify those that can grow in small spaces, demonstrating a means through which knowledge can be applied in practice to advance development using "urban gardening" techniques.

The training is not framed as one that teaches best-practice pedagogical methods to teachers per sé, but the environment teachers experience allows them to witness how an exploratory pedagogy might be implemented.[2] The human capital they acquire has multifaceted implications, potentially transforming their own pedagogy, as well as their conceptions of knowledge. By experiencing an environment in which they participate in the production of knowledge, they gain a deeper appreciation of the importance of

[2]As a result of this framing decisions, conversations with teachers suggest that the purpose of the training during the first two to three days of intensive study are often quite confusing, though this confusion dissipates by the end of the first ten-day intensive study period.

including larger numbers of their students in the process of producing, accumulating, and applying knowledge to advance processes of individual and community development. In short, their conception of knowledge and its role in society changes.

The teacher-trainer, or tutor's, role is to facilitate this process of discovery by reading the PSA texts together with the teachers, guiding discussion among the participants, and implementing the practices suggested by the text. Along the way, the tutor ensures that teachers can use increasingly precise language to describe what they are learning and the discoveries they are making through research — in other words, to back their opinions with sound logic and evidence. Appendix A presents an example anonymized script of a conversation between the tutor and a set of participants during a lesson that took place in January of 2019, demonstrating how this takes place during the course of the teacher training.

A common thread early on in the training is that participants want to assign definitions to concepts (such as "the general (or specific) properties of matter" in Appendix A). However, they quickly shift to a mode in which they seek to understand why such concepts are useful in the first place. This is a stark departure from a pedagogy that promotes rote memorization of definitions, lecture-centered pedagogy, and overly theoretical lessons devoid of a connection to practice. Once teachers are empowered to learn in this way, they develop a new-found love of discovery and learning. They are assisted by the KN tutors to recognize how the unique pedagogical experience served as a vital component to this transformation. We hypothesize that this appreciation of pedagogy will: 1) shift motivation and effort made by the teacher — and instill the same love of learning in their students — and 2) increase competency in facilitating research and discovery-based learning in the classroom.

These hypothesized changes at the level of pedagogy will naturally lead to changes within the student as well. First, as the change in pedagogy leads to an increase in student-centered learning experiences, we expect student engagement in classroom activities will increase, leading to a general increase in learning. We hypothesize that both traditional measures of student learning — such as Primary Leaving Examinations and other standardized tests — as well as higher-order, or deep, learning that connect reasoning and critical thinking skills to concrete practices that solve locally relevant problems. Second, relationships between students and teachers will become more "horizontal" and cordial. For example, teachers may be less likely to use corporeal punishment for poor behavior, a common practice in Uganda, and will increase the degree to which they converse with their students regarding matters related to their lives outside of the classroom.

Over time, both teachers and students will be able to think more creatively and clearly about how the knowledge they gain can benefit processes of community development. In other words, we expect the new approaches taken by the teachers will directly benefit the families of the students who attend treated schools and will affect the short-run aspirations and long-run choices made by the students themselves. How might this happen? As mentioned above, each of the modules in KN's training are designed to help a participant apply knowledge to learn about their own environment. The modules progressively develop the capacity of teachers to analyze their communities and become "promoters of community well-being" as a result. They conduct research by interviewing community members about food consumption, farming, hygiene, and other behaviors in the community, observing behaviors and analyzing results. As students and teachers begin to view themselves as protagonists and promoters of community well-being through research, they bring these skills to their respective households and these spillovers enable households to become more active agents of change in their own development — becoming less reliant on external support and more aware of opportunities for economic growth within their own community.

1.1 Research Questions and Hypotheses

To summarize, we articulate the below research questions and associated hypotheses as we evaluate the impacts of KN's teacher training program:

(Q1) Does the teacher training change teacher pedagogy, attitudes, and effort?

(H1.1) The teacher training will increase the following outcomes observed in classrooms:

- student engagement,
- pedagogical techniques that engage critical thinking and practical exploration,
- pedagogical techniques that facilitate understanding of concepts and deeper learning.

(H1.2) Trained teachers are intrinsically more motivated to teach and exhibit higher effort.

(H1.3) Teachers will be more willing to learn from others and collaborate with their peers following the teacher training.

(Q2) Does exposure to trained teachers improve student outcomes?

(H2.1) Traditional learning outcomes are higher for students taught by trained teachers.

- Pass-through Rate
- Primary Leaving Exam (PLE)

(H2.2) Higher order learning outcomes are higher for students taught by trained teachers.

(H2.3) Scientific capabilities are higher for students taught by trained teachers.

(H2.4) Students taught by trained teachers find more creative uses of local resources.

(H2.5) Students taught by trained teachers engage in less "risky behaviors" than their peers.

(Q3) Does the teacher training change relationships among teachers, students, and families?

(H3.1) Teachers, students and students' families will engage on a more personal level in treated schools.

(H3.2) Teachers will view themselves as more active members of the community.

(H3.3) Students will be more likely to assist one another with school-related activities.

(Q4) *Are there community effects or externalities of the program?*

(H4.1) Teachers will subsidize incomes through entrepreneurial activities.

(H4.2) Households will be more likely to engage in agricultural experimentation and innovation.

(Q5) *Does student and family exposure to treated schools affect economic outcomes?*

(H5.1) Students in treated schools are more likely to seek economic opportunities in their own villages.

(H5.2) Agricultural productivity is higher for families connected to treated schools.

We will pre-specify tests of select hypotheses in questions **(Q1)** and **(Q2)** and will reserve questions **(Q3)**, **(Q4)** and **(Q5)** for exploratory analysis in future papers.

1.2 Additional Background

KN was founded in 2007 in Jinja, Uganda with the aim of developing capacity in the local population to become protagonists of their own community development. They do so by coordinating the study of the PSA study groups in a locality. The PSA curriculum is developed through an ongoing process of action-research and has been in development since the mid-1970s. The Foundation for the Application and Teaching of the Sciences (FUNDAEC), a development organization in Colombia, South America, coordinates the process by which the PSA curriculum is developed.[3]

[3] FUNDAEC is the Spanish acronym for Fundación para la Aplicación y Enseñanza de las Ciencias. FUNDAEC aids their participants to make informed technological choices that are conducive to a healthy community life. More broadly, their aim is to develop the intellectual and social capabilities of rural populations to transform their communities. They describe some of these capabilities as: managing one's affairs responsibly with rectitude of conduct, making scientific observations and describing what they observe in the world around them, to place these observations into ever greater contexts, to build environments of unity based on an appreciation of diversity, to read with good comprehension the literature of a given field, to participate effectively in community consultation, to plan and execute business plans for small enterprises, and to treat their agricultural plots as scientists, planting and adjusting the conditions to enable their agricultural activities to meet the financial and nutritional needs of their families.

Conversations with one of the founders of FUNDAEC suggest that each book of the PSA curriculum takes at least ten years to develop through a cyclical process of action, reflection on action, study of relevant literature, and consolidation of thought into each of the modules of PSA. In the early 2000s, FUNDAEC began partnering with collaborating institutions around the world which now includes a growing network of organizations across Africa (including KN).

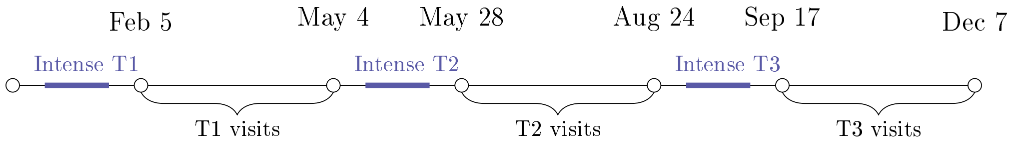
The teacher-training program was first established in 2014 as a tool for the professional development of primary school teachers within one district on the border of Central and Eastern Uganda. Given the participatory nature of the PSA curriculum, the objective of KN's teacher training program was to learn whether the study of PSA by school teachers can stimulate meaningful change in teaching practices in Uganda.

KN collaborates with district education officials to invite schools to select 2-5 teachers per school (depending on the size of the school) to participate in the in-service Teacher Training Program over the course of the school year. There are three school terms during the course of a typical school year in Uganda. In 2018, for example, term 1 took place from February 5 to May 4, term 2 from May 28 to August 24 and term 3 from September 17 to December 7. Figure 1 depicts the timeline of KN's teacher-training during a typical year. The training combines two-week periods of full-time programming with KN tutors with in-service periods during the school year.

We dub the periods with full-time programming "intense" periods. These intense periods are held over the course of a year during term breaks. Each of these periods lasts for roughly 11 days during which teachers from a set of 13 to 18 schools (total of 40 to 50 teachers) gather at a central location (typically one of the schools) from 8:00 AM to 5:00 PM to study a sequence of texts and engage in practical exercises that help link the readings with everyday reality.

During the school year, PSA tutors visit classrooms to observe how teachers are engaging with students and reflect with them on outcomes. Notwithstanding the fact that teachers are not paid to participate (only a transport subsidy is provided), training participation rates are high — roughly 80% of teachers invited to the training participate in the training

Figure 1: Timing of Teacher Training (2018)



for a majority of the sessions over the course of a year.

2 Research Strategy

2.1 Sampling

2.1.1 School Sampling Frame

We have been collaborating with KN since early 2017 on the design of a randomized roll out of their program to ensure a rigorous evaluation of their teacher training program. In the middle of 2017, we identified Budondo subcounty as the first of Jinja district's subcounties that KN would implement its program. To identify the sample in Budondo subcounty, we took a census of all the primary schools in the area which included information on the ownership status of each school, enrollment numbers, the highest class taught, and the number of teachers (among other measures). Given the project's focus on upper primary teachers, we limited schools to those that teach up to the highest primary school level, which is primary 7 in Uganda (the year that the primary leaving exam is taken). This left us with 37 remaining schools of which 2 closed doors before the end of the school year. We included the remaining 35 schools in our sample - 15 of which are government-owned. KN began working with the randomly selected treatment schools in December 2017 after baseline data was collected.

During the first two years of the study, two additional private schools in our sample closed doors because they were no longer solvent. As we were planning to expand the sample in 2019, we followed the same procedure as in 2017 in three additional subcounties in Jinja district; however, we added additional restrictions for participation to ensure schools will be relatively stable across the time-span of the study. Specifically, we limited schools to those with at least 80 students enrolled and at least 10 primary 7 students enrolled. This limited the number of schools in our sample to 67 (for a target of 100 schools total), of which 10 private schools refused to participate in the study. After randomly selecting treatment schools from the final set of 57 additional schools, KN began involving teachers in its teacher training in December 2019.

2.1.2 School Assignment to Treatment

We randomly selected 18 out of 35 Budondo primary schools and offered training opportunities only to these schools. Randomization was conducted in November 2017 by pairwise matching[4] on 7 school characteristics drawn from administrative data.[5] First, one school was chosen uniformly at random for treatment. Then, the remaining 34 schools were greedily matched into pairs according to minimum Mahalanobis distance among the 7 matching characteristics. Finally, one school from each of the 17 pairs was chosen uniformly at random for treatment. In 2019, we repeated this process with the 57 schools that joined the sample.

2.1.3 Teacher Training Attendance

Although the training offer is random, only a subset of teachers in each school (and indeed a subset of schools) take it up. Takeup need not be random. To identify the treatment effect on the treated, we asked all pre-randomization schools to nominate the teachers they would send in the event that their school is assigned to treatment. We can make valid teacher-level comparisons across treatment and control schools if we use the interaction of treatment and nomination as an instrument for takeup. However, we will not specify any such analysis in the pre-analysis plan and leave discussion of this approach to future papers.

[4] Algorithm from Bruhn & McKenzie (2009)

[5] Government ownership, student enrollment, number of teachers, fraction of teachers who teach 6th and 7th year primary school, mean population under 15 years old in feeder villages, mean number of years that teachers have worked at the school, and fraction of teachers who completed upper secondary school.

2.1.4 Teacher, Student and Household Sample

Outcomes will be measured at the teacher, student and household level to address the research questions specified above. For the analysis specified in the pre-analysis plan we will limit our analysis to the following three sampling frames: upper primary teachers, students in primary six (survey and researcher-administered assessment), students sitting for primary leaving exams. For teacher outcomes, we interview the universe of upper primary teachers in our school sample. Furthermore, we conduct classroom observations of all classes during the morning hours of operation at schools in the sample (typically between 8:00 AM and 1:00 PM, though this can vary by school and school day).

Our sampling frame for students is the universe of primary six students in each school. We select ten such students in each school to partake in our sample, balancing across school-specified performance quintiles and student gender. We collect performance quintiles every term and use the most updated performance quintile available prior to conducting interviews and assessments in term 3. We group students in quintile-gender cohorts and randomly order students within cohorts. If the first student in this cohort is unavailable for interviews on the given day of surveying and researcher-provided assessments, we draw the next individual within the cohort for our sample. For primarily leaving exam (PLE) outcomes, we will analyze the universe of outcomes from each school in our sample.

We refrain from describing the household sample in the pre-analysis plan since it will not be used to pre-specify hypothesis tests.

2.1.5 Attrition from the Sample

Two kinds of attrition are possible: attrition of schools and attrition of teachers.[6] If a school is missing from the sampling frame, we will exclude its partner school from our regressions. This leaves our estimates unbiased regardless of how attrition depends on treatment or potential outcomes. We will not specify our approach to dealing with attrition of teachers in

[6]For the pre-specified hypothesis tests, we will not be analyzing students in a panel. Rather, we will analyze students in primary schools at multiple cross-sections across time, so attrition is not an issue for student outcomes.

the pre-analysis plan.

2.2 Fieldwork

2.2.1 Instruments

Our study will make use of a dataset compiled out of household, student, and teacher surveys, classroom observations, school administrative data, and community census data from a subset of study regions. These data are collected as follows:

1. *Survey data*: Survey data is collected by trained enumerators using the survey instruments described above. There are four types of survey data: teacher, student, household, and community-leader surveys. The teacher and student surveys are conducted at the schools in our sample during the third term of a given academic year. The household and community-leader surveys take place at the homesteads of a given respondent. However, during the ongoing Covid-19 pandemic, we have been forced to conduct a subset of these surveys over the phone. The degree to which phone surveys will replace in-person surveys is yet to be determined. However, since we will not be pre-specifying tests using household surveys, we will not describe these surveys and associated observations in the pre-analysis plan.

2. *Classroom observations*: Classroom observation data is collected by trained enumerators leveraging the Stallings and TIPPS observation tools.[7] Stallings provides information on teacher time-use, while TIPPS analyzes the quality of teaching delivery. Stallings enumerators are trained to administer the classroom observation tool over a one-week long training in which they observe mock classes taught by primary school teachers in a nearby district. Observations are in the form of a "snapshot" in which the enumerator observes what the teacher is doing, the degree of student engagement in the activity facilitated by the teacher, what non-engaged students are doing,

[7] TIPPS stands for Teacher Instructional Practices and Processes System and is developed specifically for developing country contexts by a team spearheaded by Edward Seidman at NYU. It has been validated in the Ugandan primary school setting.

and the materials used by teachers and students. We calibrate the Stallings tool so that enumerators collected ten such snapshots for each scheduled class in the school's timetable.

When in the classroom, Stallings enumerators also set up tablets used to record videos of the lectures. These videos are used by a separate set of enumerators to code observations using the TIPPS tool. TIPPS requires a higher degree of skill among enumerators. Each enumerator watches video of a teacher lecture twice and proceeds to code responses related to the degree to which the teacher facilitates critical thinking, cooperation among students, deeper understanding of concepts, and more. TIPPS enumerators go through a one-week long training facilitated by a project co-director of TIPPS at NYU. During the training, their responses are compared to an "answer key" generated by the TIPPS project directors. The subset of enumerators who receive TIPPS certification will be employed in our study. In 2020, we also began to conduct random unannounced spot checks prior to the Covid-19 pandemic.

3. *School administrative data*: School administrative data are provided by all schools that agreed to participate in the study. These data include information on teacher and student attendance, student registration, and student exam scores in each term. While they are required by law to do so, many schools did not systematically collect this data on their own prior to engaging with the study so data is often incomplete, especially for earlier rounds of the study. As schools develop the habit of collecting and documenting information, the number of schools with complete administrative data increases. As described above, we will drop pairs of treatment and control schools when a school is missing data for the time-period relevant to a particular analysis.

4. *Researcher Administered Student Assessment*: Following student interviews, students participated in a researcher-administered assessment. We describe this assessment in greater detail, including the variables we construct out of it, in section 3.1.

5. *Uganda National Examination Bureau Data*: Primary 7 "candidates" register for the Primary Leaving Examination (PLE) in June of each school year. They sit for the examinations in November and results are provided in January to assess whether a candidate can attend secondary school. Students from different schools often sit at a centralized school for exam-taking purposes. Therefore, we need to collect student registration numbers from each school after students have registered but prior to students having taken the PLE. We merge these registration numbers with official records from the Uganda National Examination Bureau (UNEB) which provide student-level results for Math, English, Science and Social Studies. We purchase these results in February of each year relevant to our study. Unfortunately, we were unable to acquire baseline PLE results for the full set of schools in 2017 since our study began in November of that year after PLE examinations had been taken.[8]

6. *Budondo census data*: Finally, we collect census data in Budondo sub-county which we use to identify household-level assignment to treated school at baseline in addition to demographic and economic characteristics of each household. Census data will not be utilized in our pre-analysis plan, though we intend to use it to measure community exposure to treated schools as well as changes to household demand for treated vs. control schools over time. Census data will be used to supplement school administrative data on student registration. Using an additional dataset for student registration data allows to test and potentially increase the accuracy of registration data used.

2.2.2 Data Processing

The data will be processed by the PI team and research assistants based on this pre-analysis plan.

[8]Given poor practices around record-keeping in a subset of schools, we were unable to compile the list of pre-registered PLE takers for the full set of schools in our study.

3 Empirical Analysis

3.1 Variables

We will organize the description of measurements according to the research questions articulated in section 1.1. Each of these measures will be constructed using data from the above mentioned sources.

*3.1.1 **(Q1)** Does the teacher training change teacher pedagogy, attitudes, and effort?*

(H1.1) Classroom Pedagogy Improves.

Using the TIPPS and Stallings classroom observation tools, we will construct measures of student engagement, critical thinking pedagogy and pedagogy facilitating deeper learning and understanding of concepts. In cases in which we construct index measures, we describe the formation of the index measure in appendix section B. We construct the variables under each outcome measure for hypothesis **(H1.1)** in the following manner.

- **Student Engagement *(Stallings)*:** For each snapshot, the enumerator needs to identify the activity the teacher is engaging students in and indicate the materials the teacher is using for this activity. Then, the enumerator needs to specify a response to the question "How many pupils are engaged in the activity with the teachers?" The enumerator can respond with 1) "No pupil," 2) "One pupil," 3) "A few pupils," 4) "Half of the pupils," 5) "Most of the pupils," or 6) "All pupils."[9]

- **Critical thinking pedagogy and exploration *(TIPPS)*:** Enumerators using the TIPPS tool have a series of encoded concepts that identify whether a teacher practices exploratory pedagogy and engages students' critical thinking through their teaching style. We combine these concepts into an index using principal components

[9] We follow the instructions in World Bank Group (2015) and emphasize that even if only one student is involved in an activity, students are coded as "engaged" if they are paying attention to the activity. For example, if a single student is asked by the teacher to read a passage out loud, all students are considered "engage" if they are at least listening attentively to the passage being read.

analysis we call "Exploratory Pedagogy and Critical Thinking." The variables we use are identified in appendix section B. We will impose a standard normal distribution of this index to facilitate interpretation of results.

- **Pedagogy for Understanding Concepts and Deeper Learning *(TIPPS)*:** Again, a series of concepts in the TIPPS tool identify whether a teacher draws out student understanding through extended conversation and by scaffolding concepts in class, for example. We combine these concepts into an index using principal components analysis we call "Deep Learning Content Delivery ." The variables we use are identified in appendix section B. We will impose a standard normal distribution of this index to facilitate interpretation of results.

- **Students ask questions in class *(Teacher Outcome from Student Survey)*:** We ask students to indicate whether they asked teachers any questions during class (in terms 2 and 3) to help them with a problem they didn't understand. If they respond yes, we ask them to indicate the teachers whom they asked to explain something in this way. We transform this variable into a teacher-level observation by summing student responses at the teacher-level and dividing by the number of students surveyed. Separately, we also ask P6 students to identify which of the upper elementary teachers regularly teach them in their school. We restrict analysis of this variable to those teachers who teach P6 students as indicated by at least 90% of surveyed students mentioning a teacher in this way. Formally, and for each school, let y_{ijk} be equal to one if teacher i was mentioned by student j in school k and zero otherwise. The variable we construct is equal to $y_{ik} = \sum_j y_{ijk}/N_k$ where N_k reflects the number of students surveyed in school k.

- **Corporeal Punishment *(Teacher Outcome from Student Survey)*:** We gently ask students to indicate whether a teacher has "caned" or beaten him or her. We stress to the students that they do not have to respond to the question if they do not want to, but that their responses will remain confidential no matter how they respond.[10] If the students say "yes," we ask them to indicate which teachers caned or beat him or her. We construct teacher-level outcomes in the same manner as the "Students ask questions in class" variable above.

(H1.2) Teacher Motivation and Effort:

- **Teacher Attendance *(School Admin Data)***: Schools are required to collect daily teacher sign-in sheets, however not all schools do so. For the schools that collect this information, we sum and then digitize the days of teacher attendance per term at the teacher level.[11]

- **Teacher Knowledge of Student *(Teacher Outcome from Student Survey)***: In the student survey we ask students in P6 to indicate their attendance at the school during the prior two school days. We additionally ask these students to indicate their relationship with the people they live with. Then, in the teacher survey, we randomly selected six of the students we surveyed and added two additional names of students who do not attend the teacher's school (hereafter dubbed "fake students"). We first ask the teacher to identify whether the named student attends the school (question 1). If the teacher responds "yes" to this question, we proceed to ask, questions 2, 3 — whether the named student was present in school during the previous 1-2 school days (one question for each school day) — and question 4 — what is the relationship that the student has with the people he or she stays with. We match teacher and student responses and create an index reflecting the accuracy of the teacher's response.[12]

  Formally, and for each school, let q_{ijk} reflect teacher i's response regarding student j for question $k = \{1, 2, 3, 4\}$ — student j's response

[10] "Caning" is a common term used for using something like a thin bamboo stem to hit a child on his or her head as a form of corporeal discipline.

[11] Depending on funding, we will also try to conduct random unannounced spot checks on a regular basis from each school in the study. The Stallings Instrument also indicates whether the normal teacher is in the classroom during the specified time in the school's timetable.

[12] Separately, we also ask P6 students to identify which of the upper elementary teachers regularly teach them in their school. We restrict analysis of this variable to those teachers who teach P6 students as indicated by at least 90% of surveyed students mentioning the given teacher.

for question k is reflected by s_{jk}. The index we construct amounts to the percent of teacher i's correct responses across all students and questions:

$$\text{Teacher Knowledge of Student}_i = \frac{\sum_j \sum_k \mathbb{1}(q_{ijk} = s_{jk})}{\sum_j \sum_k 1}, \qquad (1)$$

where $\mathbb{1}(q_{ijk} = s_{jk})$ is equal to one when the teacher's response matches the student's response and zero otherwise.

(H1.3) Teacher Collaboration:

- **Teachers learn from and collaborate with colleagues (Teacher Survey):** Recall that to compile our sample for teacher interviews, we acquire a complete list of teachers in each school in the second term of a given school year and interview upper primary teachers in the third term of that school year. We pre-load teacher names in each school into our teacher survey and ask each interviewed teacher to respond to the following three questions regarding classroom and school learning/collaboration with other teachers:

  1. Have you and [*colleague*] spoken about how to better manage your classrooms?
  2. Have you or [*colleague*] visited any of each others classroom to help improve teaching practices?
  3. Have you planned classroom activities together with [*colleague*]?

  Thus, for a given teacher i and colleague j we observe i's description of the symmetric relationship she has with j — $y_{ij}^{net,k}$ for questions $k \in \{1, 2, 3\}$. Importantly, if school s has a sample of N_s teachers, then we observe both i's responses ($y_{ijs}^{net,k}$) and j's response ($y_{jis}^{net,k}$) for each ij or ji pair in $\{1, 2, \cdots N_s\}$. To add precision to our measures, we only consider affirmative responses in which ijs and jis agree.[13]

[13] A subset of sampled teachers are not included in our pre-loaded survey program because they were not listed as school teachers by school administrators prior to term 3. Therefore, for these teachers we only have observations for either $y_{ijs}^{net,k}$ or $y_{jis}^{net,k}$ but not both. We exclude these teachers (whether present in i or j) from our analysis which preserves the precision of our measure and imposes symmetry on dyadic responses.

In other words,

$$y_{ijs}^{net*,k} = y_{ijs}^{net*,k} = (y_{jis}^{net,k} = \text{Y} \ \cap \ y_{ijs}^{net,k} = \text{Y}) \\ \text{For all } k \text{ in } \{1,2,3\} \text{ and } i,\ j \text{ in } N_s \tag{2}$$

where $(y_{jis}^{net,k} = \text{Y} \ \cap \ y_{ijs}^{net,k} = \text{Y})$ indicates that both i and j answered "Yes" to social network question number k.

The final variable we construct for the pre-analysis plan is the dyad-specific average response for each question $y_{ijs}^{net,k}$. In other words, we analyze

$$y_{ijs}^{net} = \frac{1}{3} \sum_{k \in \{1,2,3\}} y_{ijs}^{net*,k} \tag{3}$$

*3.1.2 **(Q2)** Does the teacher training change student outcomes?*

(H2.1) Traditional learning outcomes increase *(PLE scores, Student Assessment and Student Survey)*:

- **Primary Leaving Examination (PLE) at the end of P7 *(UNEB and School Admin Data)*:** As described above, we collect school records on PLE registration and combine it with official data from UNEB. Lower scores indicate higher performance and range from 1 to 9 for each subject. To ease interpretation, we flip this relationship and make 9 the highest performance and 1 the lowest performance. Furthermore, we aggregate scores from each of the four subjects in a manner that reflects the official approach taken by UNEB to determine student performance. Therefore, our aggregate measure ranges from 4 (lowest performance) to 36 (highest performance) which we standardize according to the control school mean and standard deviation.

- **Student Pass-through Rates *(Student Survey)*:** In the student survey, we ask students to indicate the grade they are currently in, the grade they were in one year prior and the grade they were in two years prior. We indicate that a student has had a successful pass-through rate if the most recent year's grade is one (or more) grade ahead of the previous year's grade. Formally, and for example, a successful pass-through for student i in school s in time t (t reflects the school year) is characterized by

$$y_{ist}^{pass} = \mathbb{1}(\text{Grade}_t - 1 >= \text{Grade}_{t-1}) \tag{4}$$

where $\mathbb{1}$ reflects an indicator variable equal to one if $(\text{Grade}_t - 1 >= \text{Grade}_{t-1})$ is true and zero otherwise.[14]

- **Researcher-provided Student Assessment *(Student Assessment)*:**

(H2.2) Higher order learning outcomes increase

- **Measuring Recall, Understanding and Critical Thinking *(Student Assessment)*:** We borrow assessment questions from SIPRO (2019). SIPRO is a Ugandan company that sells exams to primary schools across the country for beginning-, mid- and end-of-term exams that cohere with the National Curriculum. Specifically, since our survey sample only includes P6 students at the beginning of term 3, we utilize a subset of the term 2 middle-of-term assessment questions sold to schools. Following Burdett (2017), we organize questions according to whether they measure recall, understanding (or application of knowledge), or reasoning/critical thinking. However, also similar to Burdett (2017), we found very few questions that address the second two categories of questions in SIPRO's standard assessment.

  To address this shortcoming, we designed several questions that intend to measure higher-order learning principles described in Liu et al. (2014). Specifically, we tried to identify one question for each of the following elements of Liu et al.'s (2014) conceptual framework for critical thinking assessments: 1) evaluate evidence and its use, 2) analyze and evaluate arguments, 3) understand implications and consequences, 4) develop sound and valid arguments, 5) understand

[14] We furthermore ask whether students attended the same school in the prior year. In the pre-analysis plan we will pre-commit to analyzing this variable for all students in our sample, though exploratory analysis will analyze whether this variable responds differently when the analysis only includes students who attended the same school year after year.

causation and explanation. These elements cohere with Burdett's (2017) description of reasoning/critical thinking.

Appendix section C displays the format we used for the student assessments. Questions 1 through 5 were borrowed directly from SIPRO (2019) while questions 6 through 9 were generated by the research team. Students fill out the assessment on paper and enumerators enter student answers into our pre-programmed tablet-based form. Enumerators were also trained to assess the completeness of sentences and grammar structure of responses in question 4. Appendix table C.1 demonstrates how we classified each assessment question according to Burdett (2017).

We utilize factor analysis to predict three coarse learning outcomes based on our approach to classifying assessment questions. The three outcomes reflect different dimensions of student ability: 1) recall, 2) understanding and 3) reasoning or critical thinking ability. Out of the 23 opportunities to receive points from the SIPRO (2019) exam, 52% fall under recall, 39% fall under understanding and 9% fall under reasoning.

The questions we add to the student assessment all fall under the reasoning category. However, since such questions are rare in the Ugandan context, we only use them to predict reasoning ability if they achieve a benchmark threshold of validity.[15] Specifically, since 2 of the questions from SIPRO (2019) represent the reasoning criteria, we analyze correlations between these two questions and each of the seven questions we construct ourselves. If the average correlation across the two questions that provide benchmark for our "reasoning" criteria is less than 0.05, we exclude the question from our construction of the reasoning variable.

(H2.3) Field-based scientific competency increases *(Science Shows)*: The Ugandan National Curriculum Development Centre (NCDC) is pushing a curriculum transition towards a "competency-based" framework

[15]Indeed, Burdett (2017) shows that such questions are very rare throughout the African continent, though Uganda utilizes such questions with greater frequency than its sister countries across Africa.

for education, starting with curriculum reform in secondary education. Recently, however, it began specifying expected competencies in students in primary training as well. Given the national emphasis on scientific competencies, and together with the Jinja District Education Office (DEO), we develop a field-based measure of testing scientific competencies in students across all of the schools in our study area. This activity is designed to measure the following competencies articulated by NCDC's (2016) curriculum guide for Primary 6 students:

- Pupils apply correct scientific processes in investigations of various phenomena by:
  - identifying problems,
  - designing and practicing scientific investigation processes,
  - examining the evidence useful in inferences,
  - demonstrating the skills of observation, classification, accurate measurement and recording,
  - making predictions and formulating hypothesis for evidence,
  - communicating findings accurately and honestly,
  - analyzing causes and effects,
  - using a variety of sources for acquiring information, and
  - recording information with reasonable accuracy.

Prior to the science show. In order to measure these competencies we worked with the DEO to organize science shows at each of the schools in our study and trained judges to utilize a rubric that measures student competencies along each of these criteria. We describe each of these activities in turn below.

First, NCDC (2016) articulates a set of activities that P6 students could implement in the course of a science show. In term two and three science includes lessons on natural resource conservation (with an emphasis on soil) and clean water preparation (for drinking) at home. Together with the DEO, we drafted a letter to each school that, in part, reads:

"Each stream of primary 6 class will organize 3 or 4 groups of students who will prepare a science-oriented activity on one of two subject competencies:

1. "Experiments that compare at least two different approaches of conserving resources in the community such as soil, minerals, fuel, water, and air.
2. "Experiments that compare at least two different approaches of preparing clean and safe water for drinking and washing."

This letter is sent to schools towards the end of term 2 (typically July) and schools are informed that science shows will take place in the middle of term 3 (October), giving them three months to prepare. The letter states that each group of students will present the results of their experiments in front of pre-appointed science show judges and each presentation should take 10-15 minutes with an expected 5 additional minutes of questions and answers with the judges.

For schools participating in the science shows, the student surveys ask students to describe 1) the research question they are exploring in their science show project, 2) the hypothesis around the question, 3) why they think the hypothesis might be true, 4) which two approaches they will be comparing in their experiment and 5), what they should observe from these approaches according to their hypothesis. The purpose for these questions was to ensure that schools were preparing adequately for the science show and also to provide the team of researchers material that could be adapted to utilize during the training of science show judges.

The rubric provided to the judges is displayed in appendix D and is shared with each school as an attachment to the letter. We measure each of NCDC's desired competencies on a ten point scale using twelve questions. Table 1 displays the rubric item used to assess each of these competencies. Our aim is to make each measure as objective a description of each competency as possible. Schools are also told that the DEO will invite a selection of community members to attend each show on the specified day and time and the highest ranking school will receive a certificate from the DEO that will also be displayed at the district office headquarters.

One week prior to the training, IPA Uganda helped identify a pool of six highly qualified enumerators to act as science show judges who participated in a training. The researchers modified student survey responses that emulated possible science shows the team would be expected to judge in the following month and asked each of the judges to provide scores for each of the items on the rubric. For three days, judges spent one hour assessing a handful of mock science shows using the form containing the rubric. Each of the judge's scores was entered into an excel spreadsheet and compared with one another. A PI for the project then identified questions where there was disagreement around the responses provided by each judge. Though imperfect, this process allowed the PIs to identify judges who could articulate the reasons behind the scores they provided in a clear and sound manner. Through this process, two of the judges were selected to work as judges for the science show on behalf of the PI and the DEO.

The day of the science show. Science shows were held in two to three schools each day from middle of October 2019 until the beginning of November 2019 (9:00 AM, 11:30 AM, and 2:30). The team of science show judges and two assistants (who also serve as backup judges) arrived at each school one hour prior to each show to assist the headteacher with logistics. In a typical science show, the group of students tasked with presenting their experiment lead the judges and community members to the site of their experiment — agricultural experiments were often implemented outdoors near the school's garden plot. Groups differed in the modality of their presentation — some selected a group leader who was tasked with presenting the work of the group, others presented their experiment as a team. Regardless of the mode of presentation, the judges were trained to ask a set of questions that they would direct to the entire team and that they would use in scoring each presentation. These question types are also articulated in the judge rubric in appendix D and are summarized by the following four items:

1. When did you start preparing for this activity? How did you come up with the question that you studied?

2. Describe how you decided to investigate the question?

3. How did you measure outcome [*specify the outcome from the experi-*

ment context]?

4. *[Instruction: Provide an alternative explanation for the observations made by the students. Ask the students how they would respond to this explanation.]*[16]

After engaging with each of the presentations, the judges invite the teachers and head teachers to address the students and community members. They then close by expressing their appreciation for the student and school's efforts and stating that the DEO will communicate the results of the shows at a later date.

After the science show. The judges provide scores to the research team and the DEO. The research team aggregates the scores of each group by taking the mean across all twelve quantitative variables in the judges rubric represented in table 1.

(H2.4) Student creativity increases *(Student Survey)*

Creativity Score *Student Survey*: We utilize a well-known tool designed by psychologists to measure student creativity. We follow the procedure implemented by Bradler et al. (2019), who utilize this measure in an experimental economic setting.

Bradler et al. (2019) measure creative performance with the unusual uses task originally developed as Guilford's alternative uses task (Guilford, 1967). In the unusual uses task, subjects are asked to name as many unique and unusual uses as they can in a limited amount of time for an item common to their subject context. In our setting, we asked students to identify unusual uses for three different items: a plastic bottle, maize plant remains, and bamboo stems. The enumerator first demonstrated how the activity works by providing some creative uses of a plastic bag and then asking the student to share additional ideas for creative uses of plastic bags. Once the students understand the nature of the activity, they are told that they have one minute to list as many creative and uncommon uses of the next item that they possibly can. Once they are told the item, the enumerator starts a timer on a tablet that sounds an alarm when 60

[16] This exercise was incorporated into the training of the science show judges prior to the science shows. The training was thorough enough that the judges developed a standard set of alternative explanations for most student experiments.

seconds have passed. In the interim, the enumerator writes each creative use of an object mentioned by a student to later record into the tablet.

We evaluate students' answers using two measures of the unusual uses task: "validity" and "originality".[17] Each valid use is valued at one point. We then categorize the answers of students according to their primary functions.[18] The originality of responses is measured by the statistical infrequency of answers according to the categories they corresponded to. Specifically, we give one additional point to a valid answer if less than 10% of participants gave the same answer and allotted two additional points to a valid answer as very original if less than 1% of subjects gave that answer.[19]

Answers that were unclear were counted as invalid. Table 2 shows the categories and percent responses for data collected in 2019for the task involving a plastic bottle. The respective tables for the tasks involving the remains of maize plants and bamboo stems can be found in the appendix (Tables E.1 and E.2 respectively).

3.2 Balance Checks

Tables 4 and 5 present balance tests for a selection of variables, including all available outcome variables, that were cleaned for analysis at the time of this writing. Unfortunately, due to funding and other constraints beyond our control, we do not have baseline variables for all outcome variables of interest for all schools. For example, we were only funded to collect student surveys in 2019, two years after the baseline data for the first set of schools was collected (as is documented in more detail in section 5). However, we have a full set of baseline variables for important datasets such as the teacher survey and classroom observations.

[17]We omit "flexibility," and "elaboration," measurement items in the original creative uses task (Guilford, 1967). Flexibility reflects the variety of a student's response and is determined by counting the number of different categories into which responses fall. Elaboration reflects the level of detail in their response. Since we utilized enumerators in drawing out responses from students, we feel that these two measures are more compromised by enumerator effects than "validity" and "originality," which complicate the use of these concepts in empirical analysis even with enumerator fixed effects.

[18]Unlike Bradler et al. (2019) we do not delineate response categories *ex ante*. Rather, we anonymize responses and ask two members of our research team to categorize the universe of responses. In case of disagreements a PI intervenes and makes the final decision.

[19]Bradler et al. (2019) use thresholds of 8% and 1% respectively.

The balance tests are broken up according to the intersection of the dataset used to generate the variable and the analyzed unit of observation. For example, we use both the teacher survey and student survey to construct the "Teacher Knowledge of Student" measure along with other student responses that generate teacher outcomes. These variables are analyzed separately from other variables extracted from the teacher questionnaire (the questionnaire is available upon request) since they utilize distinct datasets. We conduct seven specifications in this way across the two tables: 1) teacher survey outcomes using student variables (2019 only), 2) teacher survey outcomes (full sample), 3) teacher dyad outcomes (full sample), 4) classroom observation outcomes (full sample), 5) student survey outcomes (2019 only), 6) student assessment outcomes (2019 only), 7) student PLE outcomes (2017 only).[20] For each specification, we estimate the following equation (with indices changing according to the unit of observation):

$$\text{Treated}_{is} = \beta \mathbf{x_{is}} + \epsilon_{is}, \tag{5}$$

where Treated_{is} indicates the treatment status of individual/classroom snapshot i (or dyad ij) in school s, $\mathbf{x_{is}}$ is the vector of covariates tested for the given unit of observation and ϵ_{is} is the error term clustered at the school level. We wish to understand whether covariates are significantly different across treatment and control schools, the joint distribution (according to each specification) of these differences is reflected in β. In addition to presenting p values of each covariate, tables 4 and 5 present joint tests of orthogonality of these covariates against treatment status, which is reflected in the F score (and the p value associated with the F score). None of the p values associated with the F scores are significant at the 10% or 5% level. Future manuscripts will furthermore adjust for multiple hypothesis testing implicit in the multiple specifications we run to check for balance.

[20] At the time of this writing, PLE outcomes are only available for 2017. As we were collecting baseline PLE outcomes in early 2020, COVID-19 induced lockdowns impeded our access to schools which prevented us from collecting 2019 PLE registration numbers from each school.

3.3 Treatment Effects

3.3.1 Intent to Treat

We only pre-commit to estimating the intent to treat effect (ITT) by comparing outcomes across schools invited to participate in the training (treatment) and other (control) schools. Our estimating equation for the ITT is equation 6. We regress endline outcome y_{isp} for teacher or student i (or dyad ij) in school s and matched pair p surveyed by enumerator e on a dummy T_{sp} indicating treatment status of school s within its pair. We also include a vector of fixed effects depending on the characteristics of the variable selected but will always include pair fixed effects, γ_p, and generally include enumerator fixed effects when available, ζ_e.[21] The coefficient of interest β_1 is the intent to treat effect. Since treatment is assigned at the school level, we cluster standard errors at the school level.

$$y_{ispe} = \beta_0 + \beta_1 T_{sp} + \gamma_p + \zeta_e + \epsilon_{ispe} \tag{6}$$

A summary of all specifications, that links variables described in section 3.1 to a variant of equation 6, including the unit of observation and each specification's associated fixed effects is available in table 3.

4 Inference and Hypothesis Testing

To validate inference of our results, we report two sets of p values associated with the ITT in each specification. First, we report the p value derived from the use of the estimator specified in table 3.[22] That is, we will present p values that test the null that the ITT is significantly different from zero for the average school. Second, following Young (2018) we report the p value derived from randomization inference with at least 10,000 permutations of the treatment variable. That is, we will present p values that test the sharp null that the ITT is zero for all schools in the sample. These p values

[21] Variables stemming from classroom observations of teachers will also include "class" or "grade" fixed effects.

[22] Again, all p values and coefficients are estimated using standard errors clustered at the school level

have the added benefit that they allow for tests that are exact — with a distribution that is known no matter the sample size or characteristics of the error terms. Nevertheless, when drawing inference from our results we will take results of all tests into consideration.

Since we are testing multiple outcomes for a number of our hypotheses, we control the false discovery rate (FDR) across outcome measures within each hypothesis using the Benjamini-Hochberg method detailed in Benjamini & Hochberg (1995). The set of pre-specified table shells, containing both specification and randomization inference p values are presented in tables 6 through 8 for a series of (falsely) randomized treatment values. These tables omit variables whose baseline values have not been fully cleaned or coded at the time of this writing — such as administrative school data and TIPPS classroom outcomes. These tables also present the Benjamini-Hochberg corrected critical p values at the 5% level and furthermore highlight, using the $^{\pm}$ symbol, whether a given test can be rejected at either the 5% or 10% level after accounting for the FDR under the critical p values at the 5% and 10% levels respectively.

5 Project Timeline

As of the current writing, the following activities have been carried out:

2017 July Listed all schools in Budondo sub-county and acquired basic data on teachers and school characteristics.

October-November Baseline Data Collection in Budondo sub-county with an initial set of 35 schools. Collected data using the following modules:

- Teacher Survey
- Stallings Classroom Observations

Late November - Early December Randomized 35 schools into treatment and control using pairwise matching design. The partner organization began mobilizing 18 treatment schools to participate in the 2018 training.

2018 Ongoing throughout the year The research team collected administrative data on teacher attendance, student attendance, and student test scores as available.

January The partner organization hosted the first intensive training in Budondo sub-county.

February-April The research team started collecting past administrative data from schools and administered the following additional survey modules:

- Budondo sub-county household census
- Budondo sub-county community leader survey

May The partner organization hosted the second intensive training in Budondo sub-county

August-September The partner organization hosted the third intensive training in Budondo sub-county

2019 Ongoing throughout the year The research team collected administrative data on teacher attendance, student attendance, and student test scores as available.

January The partner organization hosted the first intensive training in Budondo sub-county for the second cohort of teachers from the same treatment schools.

February The research team purchased PLE data from UNEB for Jinja district for 2017 to 2019. We also commenced the process of collecting school-specific registration numbers for PLE candidates from each school in the sample.[23]

May The partner organization hosted the second intensive training in Budondo sub-county for the second cohort of teachers from the same treatment schools.

[23] This became necessary because we learned that many schools do not seat their PLE candidate at their own school. Rather, they send PLE candidates to sit for the examination at a "Seating Centre" school. This can become confusing because the third-party school will report all test scores registered to their Seating Centre as their own when, in fact, only a small subset of students who sat for the examination at the Seating Centre were educated by teachers at this school.

June The research team began preparations for the school-held science shows by coordinating with the District Education Office.

July The District Education Office sent letters to each school in Budondo sub-county informing them of the science shows to be held in October.

August-September The partner organization hosted the third intensive training in Budondo sub-county for the second cohort of teachers from the same treatment schools.

September-October Two-year follow up data collected in Budondo sub-county and baseline data collection in Mafubira and Kakira sub-counties. Modules included:

- Student Survey
- Student Assessment
- Stallings Classroom Observations - including video recordings of classrooms for subsequent TIPPS coding.
- Teacher Survey

October Science shows conducted in Budondo sub-county.

November - Early December Randomized 54 schools from Mafubira and Kakira sub-counties into treatment and control using pairwise matching design. The partner organization began mobilizing 27 treatment schools to participate in the 2020 training.

2020 January The partner organization hosted the first intensive training in Budondo sub-county for the third cohort of teachers from the same treatment schools. It also held the first intensive training in Mafubira and Kakira sub-counties for the first cohort of teachers in the 27 new treatment schools.

January Ranked performance of schools that participated in the science shows in Budondo and presented results to the District Education Office so that they can distribute prizes.

February Purchased PLE data from UNEB for the 2019 cohort of students in Jinja district.

March Schools closed due to Covid-19 pandemic.

April Canceled plans for Household Surveys.

July Reported *ITT* effects of preliminary administrative student-learning and teacher attendance outcomes to J-PAL as requested for extension of funding.

From June - Ongoing The partner organization began meeting with teachers from treatment schools in small groups to continue studying the materials and participating in community research and agriculture activities. These activities were embedded in the PSA training already but took place in the communities rather than in a central school.

early October Publish pre-analysis plan.

As of this writing Uganda is still experiencing the challenges brought on by the Covid-19 pandemic, including school closures. As a result, it is unclear how the exact timeline for future activities will be carried out. Since schools have effectively delayed study for most students by one year, we hope to extend the research project by at least one year as well. Our original intention was to effectively complete the project by the end of 2021, but now it seems more prudent to complete the project around the end of 2022. Moving forward, we hope the following activities will be carried out, pending the evolution of the ongoing pandemic and conditional on continuation of funding.

2020 October 15 Uganda opens schools for candidate classes, including P7 PLE registrants.

November Collect PLE registration numbers for candidates from schools in the study.

November-December Carry out household survey over the phone in Budondo sub-county (funding permitted).

By End of Year Code all classroom videos using the TIPPS tool.

By End of Year Publish preliminary working paper analyzing results from Budondo.

2021 Ongoing throughout the year if schools are open The research team collect administrative data on teacher attendance, student attendance, PLE registration information, and student test scores as available.

Ongoing throughout the year The partner organization will continue teacher trainings in small group settings until it is safe and legal to allow larger groups to meet.

January Uganda's President will announce plans for non-candidate class, potentially opening schools in 2021.

March-April Candidates will take PLE exam. We will purchase data as they become available.

2022 Under the assumption that the study will extend to 2022 and it will resemble a "normal" school year.

Ongoing throughout the year if schools are open The research team collect administrative data on teacher attendance, student attendance, PLE registration information, and student test scores as available.

January The partner organization hosts the first intensive training in Budondo sub-county for the fourth cohort of teachers from the same treatment schools and in Mafubira and Kakira sub-counties for the second cohort of teachers.

February The research team will purchase PLE data from UNEB for 2021 candidates.

May The partner organization will host the second intensive training in 2022 for treatment schools..

June The research team will begin preparations for the school-held science shows by coordinating with the District Education Office.

July The District Education Office will send letters for a second round of science shows in October.

August-September The partner organization will host the second intensive training in 2022 for treatment schools.

September-October Two-year follow up data collected in Budondo sub-county and baseline data collection in Mafubira and Kakira sub-counties. Modules include:

- Student Survey
- Student Assessment
- Stallings Classroom Observations - including video recordings of classrooms for subsequent TIPPS coding.
- Teacher Survey

October Science shows across all schools in the study.

November - Early December Carry out household survey in Budondo sub-county, including long-term follow up with students attending treatment and control schools in 2017 (funding permitted).

2023 February The research team will purchase PLE data from UNEB for 2022 candidates.

By June Code all classroom observation videos using the TIPPS tool and clean all administrative data.

By End of Year Publish complete working paper and submit for peer review.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

Bradler, C., Neckermann, S., & Warnke, A. J. (2019). Incentivizing Creativity: A Large-Scale Experiment with Performance Bonuses and Gifts. *Journal of Labor Economics*, *37*(3), 793–851.

Bruhn, M., & McKenzie, D. (2009, October). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, *1*(4), 200-232. Retrieved from http://www.aeaweb.org/articles?id=10.1257/app.1.4.200 doi: 10.1257/app.1.4.200

Burdett, N. (2017). Review of High Stakes Examination Instruments in Primary and Secondary School in Developing Countries. *Research on Improving Systems of Education (RISE) Working Paper 17/018*. Retrieved from https://riseprogramme.org/sites/default/files/publications/RISE_WP-018_Burdett_0.pdf

Guilford, J. P. (1967). *The Nature of Human Intelligence.* McGraw-Hill.

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing Critical Thinking in Higher Education: Current State and Directions for Next-generation Assessment. *ETS Research Report Series*, *2014*(1), 1–23.

National Curriculum Development Centre (NCDC). (2016). *Primary 6 Science Curriculum.* National Curriculum Development Centre.

Seidman, E., Kim, S., Raza, M., Ishihara, M., & Halpin, P. F. (2018). Assessment of Pedagogical Practices and Processes in Low and Middle Income Countries: Findings from Secondary School Classrooms in Uganda. *Teaching and Teacher Education*, *71*, 283–296.

SIPRO. (2019). *Sipro Mid-term II Examination 2019: Primary Six.* The SIPRO Educational Services Ltd.

Uganda Bureau of Statistics. (2018). *National Labour Frce Survey 2016/17 (Mai Report)* (Tech. Rep.). Retrieved from https://www.ubos.org/wp-content/uploads/publications/10_2018Report_national_labour_force_survey_2016_17.pdf

Uganda Bureau of Statistics. (2019). *The International Labour Day* (Tech. Rep.). Retrieved from https://www.ubos.org/wp-content/uploads/2019/05/UBOS-LABOUR-DAY-BOOK-final.pdf

World Bank Group. (2015). *User Guide: Conducting Classroom Observations.* World Bank Group.

Young, A. (2018, 11). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results.

The Quarterly Journal of Economics, *134*(2), 557-598. Retrieved from https://doi.org/10.1093/qje/qjy029 doi: 10.1093/qje/qjy029

1. Identifying the problem: The pupils clearly stated the problem they are hoping to address with their project.
1 = Did not state the problem at all.
5 = State a problem, but not clearly.
10 = Clearly stated the problem.

2. Relevance of the problem: The pupils clearly stated why the problem is an important one to address.
1 = Did not indicate why the problem is important.
5 = Indicated why the problem is important, but not convincingly.
10 = Convincingly indicated why the problem is important.

3. Designing an experiment: The pupils conducted an experiment that clearly outlines how they would study the problem they identified.
1 = Did not conduct an experiment at all.
5 = Conducted an experiment that was not clear.
10 = Conducted a very clear experiment.

4. Designing an experiment: The experiment was very creative — it tries to learn about the problem in a way that others have tried before.
1 = Others have tried to address the problem in a similar manner.
5 = The approach taken was creative, but not original.
10 = The approach was creative and original (never seen before).

5. Designing an experiment: The pupils tested a technology not seen before.
1 = The technologies tested are known by all.
5 = The technologies are known by some but not all.
10 = The technologies were completely original and creative.

6. Describing a hypothesis: The students had a clearly articulated hypothesis.
1 = I had no idea what hypothesis the students were testing.
5 = The students mentioned a hypothesis, but it was not clear.
10 = The students mentioned a very clear hypothesis, comparing at least two groups.

7. Analyzing a hypothesis: The students clearly linked their hypothesis to the design of the experiment.
1 = There was no link between the hypothesis and the experiment.
5 = The link between the hypothesis and experiment was there but not clear.
10 = It was very clear how the experiment would address the hypothesis.

8. Making observations: The students captured observations in a systematic manner (for example, by using a logbook).
1 = There is no evidence that the students recorded observations.
5 = Students recorded observations, but not in a structured way.
10 = The observations were recorded in a very structured and systematic way.

9. Making observations: The students made accurate measurements of their observations using relevant instruments.
1 = The students did not make measurements.
5 = The students made measurements that were not accurate.
10 = The students made very accurate measurements.

10. Sources of information: The students used a sufficient number of information sources to make their conclusions (for examples, they ran the experiment more than once or they measured more than one outcome.
1 = The students did not use any sources of information to support conclusions.
5 = There were at least two sources of information in support of conclusions.
10 = The students made every effort to support conclusions with all available evidence.

11. Communicating information: The students communicated their findings independently.
1 = The students could not describe what they did without help.
5 = The students described their work sometimes on their own, sometimes with help.
10 = The students were the only ones describing their work.

12. Communicating information: The students communicated their findings accurately.
1 = The students looked like they did not understand what they were saying.
5 = The students understood what they were saying, but it wasn't always connected to what they did and learned from their experiments.
10 = The students understood their work and connected it to their experiment.

Table 1: Science Show: Judge's Rubric

Category	Percent
Storage	40.1
Drinking	26.2
Toy	22.7
Irrigation	17.1
Planting	14.4
Income	11.4
Charcoal Lighter	11.1
Recycling	8.6
Construction	8.5
Funnel	4.7
Decoration	3.6
Plates	2.3
Musical Instrument	1.5
Stove Stand	1.4
Animal Feeders	1.3
Washing	1.1
Measurement	0.9
Scare Crow	0.8
Tool Handle	0.8
Jerrycan Repair	0.8
Rubbish Container	0.7
Filter	0.7
Curtains	0.6
Spray	0.6
⋮	⋮

Category	Percent
⋮	⋮
Fishing	0.6
Light Protector	0.5
Science Experiments	0.5
Shoes	0.4
Life Jacket	0.3
Plastic Wire	0.3
Workout Equipment	0.3
Weapon	0.3
Control Soil Erosion	0.3
Gutter	0.2
Pipes	0.1
Make Dice	0.1
Making Bricks	0.1
Tyre	0.1
Communication Lines	0.1
Roof	0.1

Note: This table shows the categories used for the student creativity scores for the exercise involving plastic bottles. Each mention receives one point and objects mentioned less than 10% and 1% frequency receive one and two extra points respectively.

Table 2: Categories used for creativity scores for unusual uses of plastic bottles

Table 3: Summary of All Pre-analysis Specifications

Dependent Variables:	Fixed Effects			Estimator	Datasets	Unit of Obs.	Notes
	Pair	Enum	Grade				
(H1.1) Classroom Pedagogy Improves							
Student Engagement	✓	✓	✓	Ologit	Stallings	Classroom	Ordered proportion of engaged students.
Critical thinking and exploration	✓	✓	✓	OLS	TIPPS	Classroom	Factor analysis using classroom observations
Understanding concepts and deep learning	✓	✓	✓	OLS	TIPPS	Classroom	Factor analysis using classroom observations
Students ask questions	✓	✓		Tobit	Student Survey	Teacher	Transform student response into teacher-level variable. Censored below by 0 and above by 1.
Corporeal punishment	✓	✓		Tobit	Student Survey	Teacher	Transform student response into teacher-level variable. Censored below by 0 and above by 1.
(H1.2) Teacher motivation and effort increases							

Continued on next page...

... continued from previous page

Dependent Variables:	Fixed Effects			Estimator	Datasets	Unit of Obs.	Notes
	Pair	Enum	Grade				
Teacher attendance	✓	✓		OLS	School Admin Data	Teacher	
Teacher knowledge of student	✓	✓		Tobit	Teacher and Student Survey	Teacher	Combine teacher and student responses. Censored below by 0 and above by 1.
(H1.3) Teacher collaboration increases							
Teacher learning and collaboration	✓	✓		OLS	Teacher Survey	Teacher Dyad	Mean of responses for ij pair.
(H2.1) Traditional learning outcomes increase							
English	✓			OLS	UNEB	P7 Student	Standardized score
Science	✓			OLS	UNEB	P7 Student	Standardized score
Math	✓			OLS	UNEB	P7 Student	Standardized score
Social Studies	✓			OLS	UNEB	P7 Student	Standardized score
Pass PLE	✓			OLS	UNEB	P7 Student	Standardized score
P6 Pass-through Rate	✓	✓		OLS	Student Survey	P6 Student	Most recent year
(H2.2) Higher order learning outcomes increase							

Continued on next page...

... continued from previous page

Dependent Variables:	Fixed Effects			Estimator	Datasets	Unit of Obs.	Notes
	Pair	Enum	Grade				
Understanding	✓	✓		OLS	Student Assessment	P6 Student	Standardized score of factor analysis
Critical Thinking	✓	✓		OLS	Student Assessment	P6 Student	Standardized score of factor analysis
(H2.3) Field-based scientific competency increases							
Science Show Result	✓	✓		Tobit	Science Show	Student Group	Average across all scores in rubric; censored below by 1 and above by 10
(H2.4) Student creativity increases							
Creativity score	✓	✓		Tobit	Student Survey	P6 Student	Score measuring number and uniqueness of student ideas; censored below by 0

Note: All specifications will measure the Treatment Effect using the Intent to Treat (ITT) coefficient. All standard errors across all specifications are clustered at the school level. Acronyms/Abbreviations: UNEB = "Ugandan National Examinations Bureau"; TIPPS = "Teacher Instructional Practices and Processes System"; OLS = "Ordinary Least Squares"; P7 = "Primary Seven"; P6 = "Primary Six"; Enum = "Enumerator" or "Judge"; Obs. = "Observations". Dependent variables grouped by hypotheses described in section 1.1.

Statistics	Summary Statistics								Balance Tests	
	All		Control			Treated				
	Mean	Sd	Mean	Sd	N	Mean	Sd	N	β	p value
Teacher Survey Outcomes (Student Variables — 2019 Only)										
Student Asks Questions	0.24	0.16	0.22	0.16	124	0.25	0.15	119	0.11	0.73
Corporeal Punishment	0.43	0.29	0.42	0.27	162	0.44	0.30	169	0.12	0.54
Knowledge of Student	0.45	0.23	0.43	0.21	247	0.46	0.24	238	0.48	0.12
Specification Statistics								**F Score (p value)**	1.34 (0.27)	
								Clusters	54	
Teacher Survey Outcomes										
Teacher Gender	1.39	0.49	1.40	0.49	415	1.42	0.49	422	0.03	0.50
Attended Training Last Year	0.33	0.47	0.36	0.48	415	0.32	0.47	422	-0.05	0.32
School Has Farm Land	0.65	0.48	0.67	0.47	415	0.68	0.47	422	0.01	0.94
Boys Better Pupils Than Girls?	3.33	1.52	3.33	1.52	415	3.37	1.51	422	0.01	0.72
Head Teacher Listens to Me	1.30	0.64	1.35	0.71	382	1.25	0.56	384	-0.06	0.08
How Satisfied with Job?	2.13	0.97	2.16	1.00	415	2.13	0.95	422	-0.01	0.61
Students Connect School to Family Life	1.69	0.84	1.72	0.90	415	1.68	0.78	422	-0.01	0.72
Specification Statistics								**F Score (p value)**	0.96 (0.47)	
								Clusters	89	
Teacher Dyad Outcomes										
ij Speak About Classroom Management	0.82	0.38	0.83	0.37	3,108	0.81	0.39	3,908	-0.02	0.60
ij Visit Each Others Classrooms to Learn	0.73	0.44	0.74	0.44	3,108	0.72	0.45	3,908	0.01	0.89
ij Plan Classroom Activities Together	0.62	0.49	0.64	0.48	3,108	0.59	0.49	3,908	-0.05	0.21
Specification Statistics								**F Score (p value)**	0.87 (0.46)	
								Clusters	88	
Classroom Observation Outcomes										
Share of Engaged Pupils	4.55	1.97	4.54	1.97	4,806	4.52	2.01	4,833	0.01	0.47
Activity: Q and A	0.19	0.39	0.20	0.40	4,807	0.18	0.39	4,834	-0.04	0.15
Activity: Practice and Drill	0.05	0.22	0.06	0.23	4,807	0.05	0.22	4,834	-0.04	0.34
Activity: Assignment	0.09	0.29	0.10	0.30	4,807	0.08	0.28	4,834	-0.06	0.10
Activity: Copying	0.10	0.30	0.10	0.30	4,807	0.09	0.29	4,834	-0.04	0.24
Teacher Out of Class	0.13	0.34	0.13	0.34	4,807	0.15	0.35	4,834	0.03	0.60
Materials: None	0.42	0.49	0.42	0.49	4,792	0.43	0.50	4,813	-0.03	0.65
Materials: Textbooks	0.04	0.19	0.04	0.18	4,792	0.04	0.20	4,813	0.02	0.77
Materials: Notebooks	0.10	0.30	0.11	0.31	4,792	0.10	0.30	4,813	-0.03	0.68
Materials: Blackboard	0.40	0.49	0.40	0.49	4,792	0.38	0.49	4,813	-0.03	0.63
Specification Statistics								**F Score (p value)**	0.74 (0.69)	
								Clusters	89	

Notes: This table reflects balance tests of teacher, teacher dyad, and classroom outcomes and covariates. We present summary statistics of each measure, displaying means and standard deviations for the whole sample "All," the sample of teachers/teacher dyads/classroom observations in control schools "Control," and the sample of teachers/teacher dyads/classroom observations in treatment schools "Treated" (the latter two also include number of observations). At the time of writing, we have baseline data for all schools for variables that do not use the Student Survey in their construction. For school admin data, including teacher attendance, we were in the process of collecting baseline data outside of Budondo sub-county in March 2020 when the Covid-19 related lockdowns started — we exclude these variables form balance tests here. Balance tests reflect an OLS regression with the specification $\text{Treated}_{is} = \beta X + \epsilon_{is}$ where i represents student, s represents school, X represents the vector of covariates in the rows of this table, β is the vector of coefficients associated with each covariate and ϵ_{is} is the error term clustered at the school level. Each specification is run for each data set separately, datasets are separated by the horizontal lines in the table. The F Score and number of clusters is reported for each specification. The F Score's p value (in parentheses) reports results of the null hypothesis test that coefficients are jointly orthogonal within a given specification.

Table 4: Balance Test for Teacher Outcomes and Covariates

Statistics	Summary Statistics								Balance Tests	
	All		Control			Treated				
	Mean	Sd	Mean	Sd	N	Mean	Sd	N	β	p value
Student Survey Outcomes										
Passthrough Rate 2018-2019	0.88	0.33	0.87	0.33	283	0.90	0.30	314	0.06	0.34
Passthrough Rate 2017-2018	0.91	0.29	0.91	0.29	282	0.92	0.28	314	0.03	0.68
Creativity Index	7.46	3.91	7.82	4.05	283	7.88	4.22	315	0.00	0.88
Specification Statistics								**F Score (p value)**	0.33 (0.80)	
								Clusters	54	
Student Assessment Outcomes										
Recall and Recognition	-0.10	2.32	0.46	2.04	280	0.68	2.19	317	0.01	0.65
Apply Understanding	-0.04	1.73	0.43	1.42	280	0.50	1.29	317	-0.01	0.83
Critical Thinking	-0.04	1.50	0.16	1.48	280	0.29	1.45	317	0.01	0.59
Specification Statistics								**F Score (p value)**	0.22 (0.88)	
								Clusters	54	
Student PLE Outcomes										
English	2.98	1.69	2.61	1.55	364	3.26	1.73	486	0.06	0.15
Science	3.45	1.79	3.12	1.80	364	3.69	1.75	486	0.05	0.40
Mathematics	2.92	1.51	2.75	1.50	364	3.04	1.50	486	-0.03	0.16
Social Studies	4.41	1.75	4.20	1.69	364	4.57	1.79	486	-0.06	0.33
Specification Statistics								**F Score (p value)**	1.66 (0.18)	
								Clusters	26	

Notes: This table reflects balance tests of student outcomes and covariates. We present summary statistics of each measure, displaying means and standard deviations for the whole sample "All," the sample of students in control schools "Control," and the sample of students in treatment schools "Treated" (the latter two also include number of observations). At the time of writing, we do not have baseline data for all schools. For Student Survey Outcomes and Student Assessment Outcomes, we did not collect student data until 2019, which was two years into the intervention in Budondo sub-county. For PLE outcomes, we were in the process of collecting baseline data outside of Budondo sub-county in March 2020 when the Covid-19 related lockdowns started. Balance tests reflect a regression with the specification $\text{Treated}_{is} = \beta X + \epsilon_{is}$ where i represents student, s represents school, X represents the vector of covariates in the rows of this table, β is the vector of coefficients associated with each covariate and ϵ_{is} is the error term clustered at the school level. Each specification is run for each data set separately, datasets are separated by the horizontal lines in the table. The F Score and number of clusters is reported for each specification. The F Score's p value (in parentheses) reports results of the null hypothesis test that coefficients are jointly orthogonal within a given specification.

Table 5: Balance Test for Student Outcomes and Covariates

Hypothesis:	**(H1.1): Pedagogy**		**(H1.2): Effort**		**(H1.3): Learning**
Outcome Variable:	Share of Engaged Pupils	Student Asks Questions	Corporeal Punishment	Knowledge of Student	Teacher Network
Treatment (ITT)	0.00	0.12	0.02	−0.00	0.06
	(0.08)	(0.14)	(0.09)	(0.08)	(0.06)
$\mathbf{H_0} : \boldsymbol{ITT} = \mathbf{0}$					
p value	[0.97]	[0.41]	[0.84]	[0.99]	[0.34]
RI p value	[0.99]	[0.64]	[0.90]	[1.00]	[0.51]
BH Critical p value (5%)	[0.05]	[0.03]	[0.03]	[0.05]	[0.05]
Pair FE	Yes	Yes	Yes	Yes	Yes
Enum FE	Yes	Yes	Yes	Yes	No
Grade FE	Yes	No	No	No	No
Clusters	83	83	83	83	82
Observations	6,936	314	314	314	8,380
Estimator	Ologit	Tobit	Tobit	Tobit	OLS

Notes: Standard errors are clustered at the school level. *, reflects a coefficient *p* value from the original specification, "*p value*," less than 0.1, ** less than 0.05 and *** less than 0.01. Coefficients represent the Intent to Treat effect. We report *p* values using randomization inference ("*RI p value*") as well as the Benjamini Hochberg (BH) "*BH Critical p value*" at the 5% level within hypothesis. ± suggests a significant discovery, accounting for multiple hypothesis tests, at the 10% level; ±± suggests a significant discovery at the 5% level. Units of observation in the first column are classroom snapshots observed using the Stallings instrument. Teachers in the second to fourth columns are restricted to those who at least 90% of P6 students listed as one of their teachers. Unit of observation in the fifth column is within-school teacher dyads. THIS TABLE IS A SHELL THAT USES A FALSE TREATMENT VARIABLE FOR DEMONSTRATION PURPOSES IN THE PRE-ANALYSIS PLAN.

Table 6: Teacher Outcomes for Hypotheses **(H1.1)** through **(H1.3)**: Teacher Pedagogy, Teacher Effort, Teacher Learning (TABLE SHELL USING FALSE TREATMENT VARIABLE)

Hypothesis:	(H2.1): Traditional Learning Outcomes					
Outcome Variable:	English	Science	Mathematics	Social Studies	PLE Pass	P6 Passthrough 2018-19
Treatment (*ITT*)	−0.13	0.09	0.22	−0.05	0.08	0.00
	(0.15)	(0.18)	(0.13)	(0.15)	(0.19)	(0.02)
$\mathbf{H_0}: \boldsymbol{ITT} = \mathbf{0}$						
p value	[0.38]	[0.62]	[0.10]	[0.74]	[0.68]	[0.86]
RI p value	[0.52]	[0.75]	[0.24]	[0.81]	[0.84]	[0.85]
BH Critical p value (5%)	[0.02]	[0.03]	[0.01]	[0.03]	[0.04]	[0.05]
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Enum FE	No	No	No	No	No	Yes
Clusters	29	29	29	29	29	83
Observations	899	899	899	899	964	923
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Notes: Standard errors are clustered at the school level. *, reflects a coefficient p value from the original specification, "*p value*," less than 0.1, ** less than 0.05 and *** less than 0.01. Coefficients represent the Intent to Treat effect. We report p values using randomization inference ("*RI p value*") as well as the Benjamini-Hochberg (BH) "*BH Critical p value*" at the 5% level within hypothesis. $^{\pm}$ suggests a significant discovery, accounting for multiple hypothesis tests, at the 10% level; $^{\pm\pm}$ suggests a significant discovery at the 5% level. THIS TABLE IS A SHELL THAT USES A FALSE TREATMENT VARIABLE FOR DEMONSTRATION PURPOSES IN THE PRE-ANALYSIS PLAN.

Table 7: Student Outcomes for Hypothesis **(H2.1)**: Traditional Learning Outcomes Increase (TABLE SHELL USING FALSE TREATMENT VARIABLE)

Hypotheses:	**(H2.2): Higher Order Learning**		**(H2.3): Science Show**	**(H2.4):Creativity**
Outcome Variables	Apply/Understand	Critical Thinking	Index (Mean)	Index
Treatment (*ITT*)	−0.12	−0.04	−0.52	0.16
	(0.07)	(0.07)	(0.40)	(0.19)
$\mathbf{H_0}$: $\boldsymbol{ITT} = \mathbf{0}$				
p value	[0.12]	[0.56]	[0.19]	[0.41]
RI p value	[0.30]	[0.67]	[0.44]	[0.61]
BH Critical p value (5%)	[0.03]	[0.05]	[0.05]	[0.05]
Pair FE	Yes	Yes	Yes	Yes
Enum FE	Yes	Yes	Yes	Yes
Clusters	83	83	29	83
Observations	923	923	158	925
Estimator	OLS	OLS	Tobit	Tobit

Notes: Standard errors are clustered at the school level. *, reflects a coefficient p value from the original specification, "*p value*," less than 0.1, ** less than 0.05 and *** less than 0.01. Coefficients represent the Intent to Treat effect. We report p values using randomization inference ("*RI p value*") as well as the Benjamini Hochberg (BH) "*BH Critical p value*" at the 5% level within hypothesis. $^{\pm}$ suggests a significant discovery, accounting for multiple hypothesis tests, at the 10% level; $^{\pm\pm}$ suggests a significant discovery at the 5% level. THIS TABLE IS A SHELL THAT USES A FALSE TREATMENT VARIABLE FOR DEMONSTRATION PURPOSES IN THE PRE-ANALYSIS PLAN.

Table 8: Student Outcomes for Hypotheses **(H2.2)** through **(H2.4)**: Higher Order Learning, Science Shows and Creativity (TABLE SHELL USING FALSE TREATMENT VARIABLE)

Appendix

A Example of Facilitation Style in Teacher Training

In one module, titled "Properties," the participants gain insights into the importance of precision of language in scientific thinking. After studying different parameters used to describe reality (e.g., temperature, color, size, etc.), teachers are asked to reflect on the concept of shape, one such parameter. In an exercise, the tutor guides a conversation aimed at responding to the veracity of the following statement: "Shape is a specific (not general) property of substances." Here is a snippet of the ensuing conversation recorded by one of the PIs:

Participant A It's false.

Tutor Can you go ahead and explain a little more why you think it's false?

Participant A As we were listing the specific properties of matter, we were listing shape and size. The general properties are common. Shape is one that is common to many substances. It seems every type of matter has shape. However, those specific properties that allow us to distinguish one substance from another, shape is not a type of property.

Tutor Hmm. Ok. How about others, what do you say?

Participant B For me, I say it's specific. If you want to distinguish a table and something else then you use its shape.

Participant C Let's say there's a table that has triangular shape. Does it mean that tables only have triangular shapes? What about oranges and apples? They have the same shape. But they're different substances.

Tutor It's quite challenging, isn't it? If you think about it, it's challenging to state the reason why you can't use shape or size to distinguish one thing or another. However, the other participant is saying I can use shape to distinguish one object from another. I can say that an orange is spherical and it helps me distinguish it from something that is not spherical.

Participant C That's true, but we're examining general vs. specific properties. Her statement is right, but we're trying to talk about those properties that are specific. When you talk about specific properties, they need to be such that they make the substance different from others.

Tutor [*sensing that participant C has grasped the concept, but others have not*] Is shape a general property of matter?

Participant D How do you distinguish a bus from the lorry? Don't you use shape?

Participant E We're not saying we can't describe objects using general properties. But there are those properties that further describe matter but they create a sense of difference.

At this point, the tutor points the participants back to the use of the term specific and general properties in the text. It seems many in the group have reached the understanding of participants E and C, but a few are still holding out.

Tutor Shape helps us describe substances, but is it a specific property of substances? We're not here for debating with anyone. We're here to explore concepts for the purposes of understanding. If you have reservations about what you are discussing and it appears that the majority is on the other side which you are not yet it's best to put a star so that you come back and try to understand. You shouldn't just accept that the majority is right. You have to understand on your own. Let's move on.

B Teacher Instructional Practices and Processes System (TIPPS)

The TIPPS system observes nineteen key concepts of teacher practices and classroom processes that influence pupils' cognitive and social-emotional development. The observer analyzes the extent to which each concept is implemented by a teacher in the classroom setting. The concepts describe whether the teacher:

1. creates opportunities for cooperative learning,
2. incorporates pupil ideas and interest to inform class activities and assignments,
3. uses instructional strategies to aid pupils in critical thinking,
4. uses instructional materials that facilitate learning,
5. connects activities and subject matter to a central instructional concept or learning objective,
6. connects pupils' studies to their everyday life experiences,
7. provides pupils with specific feedback to facilitate learning rather than just getting the correct answer or finishing an activity,
8. models quality language expression to advance pupil understanding and use of language,
9. asks open-ended questions and close-ended questions to facilitate deeper learning,
10. extends pupil responses to promote deeper understanding and learning of a concept,
11. uses scaffolding to promote pupils learning and understanding of subject matter,
12. creates a positive environment between the teacher and pupils and amongst peers,
13. monitors and is responsive to pupils' academic and emotional needs,
14. creates a negative environment between the teacher and pupils and amongst peers,
15. uses an encouraging tone of voice,

16. employs behavior management to create an environment that is conducive to learning,
17. establishes classroom routines to create an environment that is conducive to learning,
18. does not show favoritism towards some pupils over others, and
19. is able to engage students in learning activities.

The observer views a twenty minute video segment of a teacher's classroom session and determines the degree to which each of the nineteen statements, or its negative, describes the teacher's pedagogy. After the initial binary choice is made, the observer then indicates whether the statement (or its negative) is "Somewhat Accurate" or "Very Accurate."

Each observer completes a week long training in which they explore each concept at depth and analyze various examples in which the concept might be ambiguously interpreted (training materials available upon request). Table B.1 demonstrates how each of the concepts might be engaged with during a training session. Observers have to meet a callibration criteria if they are able to match a majority of the answers provided by the trainers for a set of example classroom observations videotaped ahead of time. The criteria and the process by which observers qualify to engage in TIPPS observations is described in Seidman et al. (2018). The observer's sheet is demonstrated in figure B.1.

To discipline the analysis of the TIPPS tool, we follow Seidman et al. (2018) and proceed by using factor analysis in order to reduce the dimensionality of the measurements to meaningful constructs. We pre-specify the constructs according to table B.2 in which a ✓indicates the use of the TIPPS concept (numbered in the description above) to construct one of five factors (in the column of the table): 1) sensitive and connected teaching, 2) exploratory pedagogy and critical thinking, 3) student expression and understanding, 4) deep learning content delivery, and 5) positive classroom management.

Concept: Teacher creates opportunities for cooperative learning.

Importance of Concept: By creating opportunities for pupils to work collaboratively, teachers enable pupils to learn from one another in a meaningful way and promote positive interactions between the pupils. These types of activities should not be confused with whole-class instruction or discussion.

Indicators Include

- Teacher structures some form of team work around the assignment or task.
- Teacher uses different types of grouping activities, such as pairing and small groups.
- Teacher monitors groupings and provides assistance to the pupils where needed.
- Teacher engages pupils in role-play or games that involve them accomplishing a goal together.

SPECIAL NOTE: Pupils may talk with one another on their own at times. However, this does not quality as teacher initiated cooperative learning.

OBSERVER SELECTION:

A: Teacher does not create opportunities for cooperative learning		B: Teacher creates opportunities for cooperative learning	
Very Accurate	Somewhat Accurate	Somewhat Accurate	Very Accurate
The teacher does not break the class into groups or pairs to work on the lesson.	The teacher breaks the class into groups or pairs, but does not define task nor teamwork responsibilities.	The teacher breaks the class into groups or pairs. The teacher does act (but not always) as a facilitator, monitoring group progress, promoting both task and teamwork.	The teacher breaks the class into groups or pairs. The teacher is acting as a facilitator, monitoring group progress, promoting both task and teamwork.
Teacher has pupils in group seating arrangement, but there is no interaction with one another.	Teacher places flashcards in the middle and tells pupils to share the flashcards with their groups. No clear instructions are provided, pupils are lost about what to do with the cards.	Teacher passes out flash cards and tells pupils to share with and quiz one another. Teacher sits back and occasionally calls out instructions if a group gets stuck or stops the activity.	Teacher passes out flash cards to groups and provides instructions and shows an example of how they should quiz one another. The teacher circulates around to check that pupils are working well together and helps when pupils get stuck.

Table B.1: TIPPS Materials — Description of Concept #1 and Examples

	Sensitive and Connected Teaching	Exploratory Pedagogy and Critical Thinking	Student Expression and Understanding	Deep Learning Content Delivery	Positive Classroom Management
TIPPS Concept				:	
1		✓			
2	✓		✓		
3		✓			
4		✓			
5				✓	
6	✓		✓		
7			✓		
8			✓		
9		✓	✓		
10			✓		
11				✓	
12	✓				✓
13	✓				
14	✓				✓
15					
16					✓
17					✓
18	✓				✓
19	✓			✓	

Note: Checkmarks, ✓, indicate that variables constructed using enumerator responses for the row-concept will be included in predicting the factor that loads on to the each variable.

Table B.2: Pre-specified Factors Loading on to TIPPS Observations

START TIME: **CLASSROOM OBSERVATION** **END TIME:**

Step One: For each item, read statements in Column A and B. Choose the statement better represents what you observed in the classroom

Step Two: For the chosen statement, check/tick (✓) either "very accurate" or "somewhat accurate" to match your observation of the item in the classroom.

Very Accurate	Somewhat Accurate	COLUMN A		COLUMN B	Somewhat Accurate	Very Accurate
v	s	Teacher does not create opportunities for cooperating learning.	**1**	Teacher creates opportunities for cooperative learning.	s	v
v	s	Teacher does not incorporate pupils' ideas and interests to inform for class activities and assignments.	**2**	Teacher incorporates pupil ideas and interests to inform class activities and assignments.	s	v
v	s	Teacher does not use instructional strategies to aid pupils in critical thinking.	**3**	Teacher uses instructional strategies to aid pupils in critical thinking.	s	v
v	s	Teacher does not use instructional materials that further learning.	**4**	Teacher uses instructional materials that facilitate learning.	s	v
v	s	Teacher does not connect activities and subject matter to a central instructional concept or learning objective.	**5**	Teacher connects activities and subject matter to a central instructional concept or learning objective.	s	v
v	s	Teacher does not connect pupils' studies to their everyday life experiences, showing the relevance of lessons outside the classroom.	**6**	Teacher connects pupils' studies to their everyday life experiences, showing the relevance of lessons outside the classroom.	s	v
v	s	Teacher does not provide pupils with specific feedback to facilitate learning rather than just getting the correct answer or finishing an activity.	**7**	Teacher provides pupils with specific feedback to facilitate learning rather than just getting the correct answer or finishing an activity.	s	v
v	s	Teacher does not model quality language expression to advance pupil understanding and use of language.	**8**	Teacher models quality language expression to advance pupil understanding and use of language.	s	v
v	s	Teacher does not ask open-ended questions or even closed-ended questions to facilitate deeper learning.	**9**	Teacher asks open-ended questions and closed-ended questions to facilitate deeper learning.	s	v
v	s	Teacher does not extend pupil responses to promote deeper understanding and learning of a concept.	**10**	Teacher extends pupil responses to promotes deeper understanding and learning of a concept.	s	v
v	s	Teacher does not use scaffolding to promote pupils learning and understanding of subject matter.	**11**	Teacher uses scaffolding to promote pupils learning and understanding of subject matter.	s	v
v	s	There are no behavioral indications of positive environment between teacher and pupils and amongst peers.	**12**	There are behavioral indications of positive environment between teacher and pupils and amongst peers.	s	v
v	s	Teacher does not monitor and is not responsive to pupils' academic and emotional needs.	**13**	Teacher monitors and is responsive to pupils' academic and emotional needs.	s	v
v	s	There are no behavioral indications of negative environment between the teacher and pupils and amongst peers.	**14**	There are behavioral indications of negative environment between the teacher and pupils and amongst peers.	s	v
v	s	Teacher tone of voice discourages pupils.	**15**	Teacher tone of voice encourages pupils.	s	v
v	s	Teacher does not employ behavior management to create an environment that is conducive to learning.	**16**	Teacher employs behavior management to create an environment that is conducive to learning.	s	v
v	s	Teacher does not establish classroom routines to create an environment that is conducive to learning.	**17**	Teacher establishes classroom routines to create an environment that is conducive to learning.	s	v
v	s	Teacher shows favoritism towards some pupils over others.	**18**	Teacher does not show favoritism toward some pupils over others.	s	v
v	s	Pupils are not engaged in learning activities.	**19**	Pupils are engaged in learning activities.	s	v

Primary Back Page

Figure B.1: Observer's Sheet for Each TIPPS Classroom Observation

C Researcher Provided Student Assessment

The following pages display the student assessment as received by the students following enumerator administration of the student survey. Each individual students participating in the survey sits independently on a bench nearby the school. Enumerators were encouraged to tell the students to "try their best" if they had any questions regarding the survey, though such questions were minimized during the piloting phase of the student assessment.

Table C.1 reflects how each of the questions in the assessment, and the modes of student response, were classified to reflect three categories of learning: 1) student recall and recognition of information (recall), 2) student's ability to understand and apply a concept (understand), 3) students ability to analyze, and evaluate an argument (reason). Throughout, we roughly follow Burdett (2017) when analyzing each question type and classifying them for our own analysis.

	Broadest Classification of Assessment Questions:				*Level 1 - Recall*		*Level 2 - Understand/Apply*									*Level 3 - Reason*							
	Second level of classification:				**Remember**		**Understand**							**Apply**		**Analyze**			**Evaluate**		**Create**		
	Third level of classification:				Recognizing	Recall	Interpreting	Exemplifying	Classifying	Summarizing	Inferring	Comparing	Explaining	Executing	Implementing	Differentiating	Organizing	Attributing	Checking	Critiquing	Generating	Planning	Producing
Question	Text	SIPRO (2019)?	Subject	Classification																			
1.a)	Give any two examples of leguminous crops.	✓	Science	Recall	✓	✓																	
1.b)	Mention any two ways of caring for crops in the garden	✓	Science	Recall	✓	✓																	
2.a)	Name two examples of vitamin deficiency diseases	✓	Science	Recall	✓	✓																	
3.a)	Which part of a plant carries out reproduction?	✓	Science	Recall	✓	✓			✓														
3.b)	Which part of a plant provides attachment to branches?	✓	Science	Recall	✓	✓			✓														
4.a)	At which school was the fire outbreak?	✓	English	Understand	✓		✓				✓												
4.a)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
4.b)	On which day did the fire break out?	✓	English	Understand	✓		✓				✓												
4.b)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
4.c)	Whose dormitory got burnt?	✓	English	Understand	✓		✓				✓												
4.c)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
4.d)	Where were the children when the fire broke out?	✓	English	Understand	✓		✓				✓												
4.d)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
4.e)	Why was the police not able to save the property?	✓	English	Understand	✓		✓			✓	✓		✓										
4.e)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
4.f)	What did the fire brigade officers tell the school management?	✓	English	Understand	✓		✓				✓												
4.f)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
4.g)	Where was the school advised to put smoke detectors?	✓	English	Understand	✓		✓				✓												
4.g)	↑Proper Sentence Structure	✓	English	Recall	✓									✓									
5.a)	How many patients were admitted on Sunday?	✓	Math	Understand	✓		✓																
5.b)	Which day had the least number of patients admitted?	✓	Math	Understand	✓		✓		✓		✓	✓	✓				✓						
5.c)	How many more patients were admitted on Wednesday than Thursday?	✓	Math	Reason	✓		✓		✓		✓	✓				✓	✓	✓					
5.d)	Find the total number of patients admitted in the whole week.	✓	Math	Reason	✓					✓							✓	✓					
6.a)	Which statement best supports Sarah's thoughts?		Critical Thinking	Reason	✓		✓				✓					✓	✓	✓					
6.b)	Which statement best supports Aminah's thoughts?		Critical Thinking	Reason	✓		✓				✓					✓	✓	✓					
6.c)	Which statement best supports both Aminah and Sarah's thoughts at the same time?		Critical Thinking	Reason	✓		✓				✓					✓	✓	✓					
6.d)	Which statement does not support either Aminah or Sarah's thoughts?		Critical Thinking	Reason	✓		✓				✓					✓	✓	✓					
7)	Sarah's Evaluation of a statement		Critical Thinking	Reason	✓		✓									✓	✓	✓	✓	✓			
8)	Aminah's Causal Reasoning		Critical Thinking	Reason	✓		✓									✓	✓	✓	✓	✓			
9)	Correlation vs. Causation		Critical Thinking	Reason	✓		✓									✓	✓	✓	✓	✓			

Table C.1: Classification of Assessment Questions

School:____

Name:____

Class:_______________________________________ Stream:___________

Age: ___________________ Sex: (Male OR Female – Circle one)

Village Where You live: ____

Name of parent/guardian: _______________________________________

Your position in class at the end of term 2: ___________________________

---

1. A. Give any two examples of leguminous crops.

   i. ____

   ii. ____

   B. Mention any two ways of caring for crops in the gardens.

   i. ____

   ii. ____

2. A. Name two examples of vitamin deficiency diseases.

   i. ____

   ii. ____

3. A. Which part of a plant carries out the following activities?

   i) reproduction. ____

   ii) provides attachment to branches.____

   B. State any two uses of leaves to plants.

   i. ____

   ii. ____

4. **Read the passage below carefully and then answer, in FULL SENTENCES, the questions that follow.**

It was a sad Saturday morning at Mogga Primary School last term. That morning shocked everyone with the sudden burning of one of the dormitories for Primary Four children. Fortunately, by the time of the fire outbreak, all of them were in class writing weekly tests, so no children were injured but a lot of property was burnt. The Police Fire Brigade rushed to the scene to save property but failed because by the time it arrived, most properties had burnt to ashes.

The fire brigade officers advised the school management to put up enough fire fighting equipment to use in case of a fire outbreak. The burnt dormitory had only one fire extinguisher which made it hard for people to stop the fire. The school did not have smoke detectors so they were advised to put smoke detectors on every building in the school.

a) At which school was the fire outbreak?

____

b) On which day did the fire break out?

____

c) Whose dormitory got burnt?

____

d) Where were the children when the fire broke out?

____

e) Why was the police not able to save the property?

____

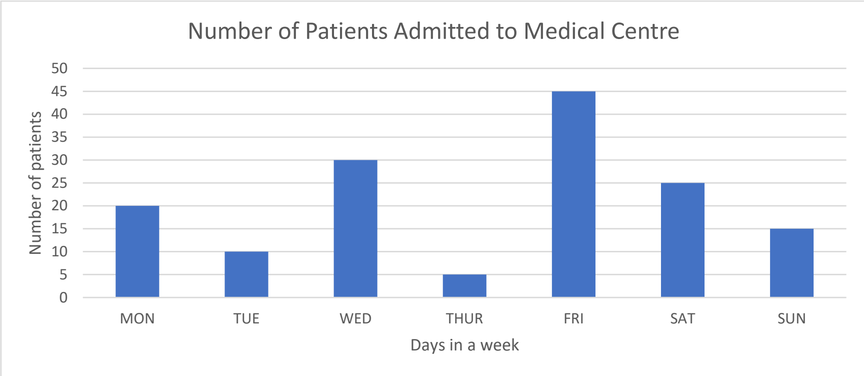
f) What did the fire brigade officers tell the school management?

____

g) Where was the school advised to put smoke detectors?

____

5. The bar graph below shows the number of patients admitted to the life care medical centre in a week. Study and use it to answer the questions that follow.



a) How many patients were admitted on Sunday?

____

b) Which day had the least number of patients admitted?

____

c) How many more patients were admitted on Wednesday than Thursday?

____

d) Find the total number of patients admitted in the whole week.

____

Question 6: In the survey, you shared your thoughts about Sarah and Aminah and how they think about education. Remember,

- **SARAH BELIEVES:** the purpose of education is to help people get jobs.
- **AMINAH BELIEVES** that the purpose of education is to help people make positive changes in their communities.

Think about these sentences.

1. 1 out of 5 (20%) of students repeat a class each year.
2. It is easier for someone with education to find a job.
3. Without education, it is difficult to improve your life.
4. In every village, the most helpful people are those with the most education.

According to you, which sentences belong in each of the blank spaces below?

(Choose one sentence for each blank space and write the sentence number in the blank)

a.) Which statement best supports Sarah's thoughts? ____

b.) Which statement best supports Aminah's thoughts? _______________________________________

c.) Which statement best supports both Aminah and Sarah's thoughts at the same time? ____________

d.) Which statement does not support either Aminah or Sarah's thoughts? _______________________

Question 7: SARAH heard that the following statement is TRUE.

1. People who are educated have more skills **BECAUSE** of their education.

SARAH decides that this also means:

2. People with skills are more educated **BECAUSE** of their skills. (CIRCLE ONE)

Is **SARAH** correct in making this decision?

CORRECT NOT CORRECT

Question 8: Aminah believes the following two statements are TRUE.

1. People who are educated serve their community **BECAUSE** of their education.
2. People who serve their community create peaceful communities **BECAUSE** of their service.

Thinking only about **AMINAH**'s beliefs, would **AMINAH** say the following:

3. People create peaceful communities **BECAUSE** of their education. (CIRCLE ONE)

YES NO

Question 9:

In many parts of the world, the following statement is true:

1. Schools that have poor performance receive more teachers to help students perform.

Is statement number 2 bellow CORRECT or NOT CORRECT?

2. If the number of teachers in a school decreases, the school's performance will improve.

CORRECT NOT CORRECT

D Judging the Science Shows

The following pages display the rubric provided to science show judges for the purpose of assessing student performance in the science shows.

School:___________________________________Group Number:_________________

Date:____________________ Time: ________________ Name of Judge: _____________________

JUDGE SUMMARIZE THE GROUP'S ACTIVITY:

WHAT ARE THE NAMES OF THE PUPILS WHO PRESENTED FOR THIS GROUP:

WHICH PUPIL SEEMS TO BE THE LEADER OF THE GROUP:

____

WHAT WAS THEIR RESEARCH QUESTION?

WHAT PROCESS ARE THEY STUDYING?

DESCRIBE THE EXPERIMENT THAT THE PUPILS ARE UNDERTAKING:

WHAT WAS THE GROUP'S HYPOTHESIS:

WHAT DID THE PUPILS OBSERVE:

School:___________________________________Group Number:__________________

Date:____________________ Time: ________________ Name of Judge: _____________________

HOW DID THEY MEASURE THESE OBSERVATIONS:

____

WHAT CONCLUSIONS DID THE PUPILS REACH:

____

____

JUDGE TALLIES:

1. Identifying the problem: The pupils clearly stated the problem they are hoping to address with their project. (Choose a number between 1 and 10)
   1 = Did not State the Problem at all;
   5 = Stated a problem, but not clearly;
   10 = Clearly stated the problem

   SCORE:

2. Relevance of the problem: The pupils clearly stated why the problem is an important one to address.
   1 = Did not indicate why the problem is important to address
   5 = Indicated why the problem is important, but not convincingly
   10 = Convincingly indicated why the problem is important

   SCORE:

3. Designing an experiment: The pupils conducted an experiment that clearly outlines how they would study the problem they identified.
   1 = Did not conduct an experiment at all;
   5 = Conducted an experiment that was not clear;
   10 = Conducted a very clear experiment;

   SCORE:

4. Designing an experiment: The experiment was very creative – it tries to learn about the problem in a way that others have tried before.
   1 = Others have tried to address the problem in a similar manner.
   5 = The approach taken was creative, but not original.
   10 = The approach was creative and original (never seen before)

   SCORE:

5. Designing an experiment: The pupils tested a technology not seen before.
   1 = The technologies tested are known by all.
   5 = The technologies are known by some but not all.
   10 = The technologies were completely original and creative.

   SCORE:

6. Describing a hypothesis: The students had a clearly articulated hypothesis.
   1 = I had no idea what hypothesis the students were testing.
   5 = The students mentioned a hypothesis, but it was not clear.
   10 = The students mentioned a very clear hypothesis, comparing at least two groups.

   SCORE:

School:___________________________________Group Number:__________________

Date:____________________ Time: ________________ Name of Judge: _____________________

7. Analyzing a hypothesis: The students clearly linked their hypothesis to the design of the experiment.
   1 = There was no link between the hypothesis and the experiment.
   5 = The link between the hypothesis and experiment was there but not clear.
   10 = It was very clear how the experiment would address the hypothesis.

   SCORE:

8. Making observations: The students captured observations in a systematic manner (for example, by using a logbook).
   1 = There is no evidence that the students recorded observations.
   5 = Students recorded observations, but not in a structured way.
   10 = The observations were recorded in a very structured and systematic way.

   SCORE:

9. Making observations: The students made accurate measurements of their observations using relevant instruments.
   1 = the students did not make measurements.
   5 = the students made measurements that were not accurate.
   10 = the students made very accurate measurements.

   SCORE:

10. Sources of information: The students used a sufficient number of information sources to make their conclusions (for example, they ran the experiment more than once or they measured more than one outcome).
    1 = The students did not use any sources of information to support conclusions.
    5 = There were at least two sources of information in support of conclusions.
    10 = The students made every effort to support conclusions with all available evidence.

    SCORE:

11. Communicating information: The students communicated their findings independently.
    1 = The students could not describe what they did without help.
    5 = The students described their work sometimes on their own, sometimes with help.
    10 = The students were the only ones describing their work.

    SCORE:

12. Communicating information: The students communicated their findings accurately.
    1 = The students looked like they did not understand what they were saying.
    5 = The students understood what they were saying, but it wasn't always connected to what they did and learned from their experiments.
    10 = The students understood their work and connected it to their experiment.

    SCORE:

13. Did the pupils demonstrate their activity in a practical way? (YES ------------ NO)

    IF YES, DESCRIBE ____

School:___________________________________Group Number:__________________

Date:____________________ Time: ________________ Name of Judge: _____________________

Q&A Session

INSTRUCTIONS TO JUDGES: DURING THE Q&A SESSION, ENCOURAGE RESPONSES FROM ALL OF THE STUDENTS. IN OTHER WORDS, IF ONLY ONE STUDENT RESPONDS TO THE QUESTION, PLEASE CALL ON OTHER STUDENTS AT RANDOM FROM WITHIN THE GROUP.

1. **JUDGE ASK THE STUDENTS: When did you start preparing for this activity? How did you come up with the question that you studied?**

Summarize Response:

INSTRUCTION TO JUDGE: Allow the students to freely respond, and then indicate which of the following best resembles the student response. [CHOOSE ONE]

A. The learners formulated their own questions or hypothesis to be tested.
B. Teacher suggests topic areas or provides samples to help learners formulate own questions or hypothesis.
C. Teacher offers learners lists of questions or hypotheses from which to select.
D. Teacher provides learners with specific stated questions or hypotheses to be investigated.
E. Other ____
F. Cannot respond because no clear choice observed.

2. **JUDGE ASK THE STUDENTS: Describe how you decided to investigate the question?**

Summarize Response:

INSTRUCTION TO JUDGE: The students shared in their presentation how they answered the question. What we want to know here is how they decided whether this approach to answering the question would be a good one.

A. Learners developed the procedures and protocols to independently plan and conduct a full investigation.
B. Teachers encouraged the learners to plan and conduct a full investigation, providing support along the way.
C. Teacher provided the guidelines for learners to plan and conduct part of an investigation. Some choices are made by learners.
D. Teacher provides the procedures and protocols for the students to conduct the investigation.
E. Other____
F. Cannot respond because no clear choice observed.

School:___________________________________Group Number:_________________

Date:____________________ Time: ________________ Name of Judge: _____________________

3. **JUDGE ASK THE STUDENTS: How did you measure outcome XXX?**

Judge's Outcome identified:

Pupil's Response:

---

4. **JUDGE ASK THE STUDENTS: Provide an alternative explanation for the observations made by the students. Ask the students how they would respond to this observation.**

Judge's Alternative Explanation:

Pupils' Response:

INSTRUCTION TO JUDGE: Select one of the following.

A. The pupils were very convincing in their response and responded to this observation adequately.
B. The pupils provided a response to this observation and it was somewhat convincing.
C. The pupils provided a response to this observation but it was not convincing.
D. The pupils did not respond at all to the alternative explanation.
E. Other____

Other Questions:

- What is the importance of this activity (you could add if they are confused about this question…. to the community)?
- ____
- ____

E Categories of creativity score

Category	Percent
Fire	41.8
Mulching	38.3
Fertilizer	30.3
Cooking	29.4
Animal Feed	28.9
Construction	10.2
Toy	6.1
Fencing	3.9
Roofing	3.8
Disciplinary Tool	2.3
Income	1.5
Cover	1.2
Scare Crow	1.1
Plant Support	1.0
Cleaning	0.8
Decoration	0.8
Hygiene	0.7
Making Ashes	0.7
Mats	0.7
Medicine	0.7
Straw Bed for Animals	0.6
Charcoal	0.6
Making Tools	0.5
Furniture	0.3
Weapon	0.3
⋮	⋮

Category	Percent
⋮	⋮
Baskets	0.2
Traditional Practices	0.2
First Aid	0.2
Wind Breaker	0.2
Building Shades	0.2
Creating Salt	0.2
Creating Smoke	0.2
Sugar Cane	0.1
Goal Posts	0.1
Riding Aid	0.1
Counters	0.1
Hat	0.1
Harvest Beans	0.1
Regerminate	0.1
Maize Packaging	0.1
Saw Dust	0.1

Note: This table shows the categories used for the student creativity scores for the exercise involving maize residual. Each mention receives one point and objects mentioned less than 10% and 1% frequency receive one and two extra points respectively.

Table E.1: Categories used for creativity scores for unusual uses of maize plant residues

Category	Percent
Construction	60.4
Disciplinary Tool	35.4
Fire Wood	20.2
Fencing	11.7
Cooking	11.3
Furniture	6.4
Income	6.1
Music Instrument	4.9
Pole	4.9
Animal Feed	4.6
Decoration	4.2
Roofing	3.8
Fishing	2.8
Plant Support	2.7
Crafts	2.5
Weapons	2.3
Medicine	2.2
Dishes	1.9
Boats	1.7
Toy	1.3
Art	1.3
Praying	0.9
Mulching	0.9
Gates	0.8
Measurement	0.6
⋮	⋮

Category	Percent
⋮	⋮
Fertilizer	0.6
Charcoal	0.6
Plumbing	0.6
Counting	0.6
Rain System	0.5
Straws	0.5
Control Soil Erosion	0.4
Wiring	0.3
Ladder	0.3
Bikes	0.2
Power Generation	0.1
Charcoal Lighter	0.1
Toothbrush/Floss	0.1
Smoking Pipe	0.1
Floaters	0.1
Pounding Machines	0.1

Note: This table shows the categories used for the student creativity scores for the exercise involving bamboo stems. Each mention receives one point and objects mentioned less than 10% and 1% frequency receive one and two extra points respectively.

Table E.2: Categories used for creativity scores for unusual uses of bamboo stems