Integrating Educational Technology with Structured Pedagogy to Improve Learning Outcomes for Every Student

Phase 1, TEACH Observations Analysis Plan

Erik T.J. Andersen, Simon Graffy, Jason T. Kerwin, and Monica Lambon-Quayefio

June 24, 2026

In this document, we describe the planned analysis for the lesson observation videos we took during waves 1 and 2 of this study. We will mark up the lesson observations using the World Bank's TEACH Primary tool to measure lesson quality. All analyses in this document are considered secondary analyses for the main study.

Sample:

The analyses in this document will be implemented using the same sample of students in the rest of this study: students who started BS1 in the 2024-25 academic year. Only those schools for which we were able to get a video will be included. We will use lessons from BS1 classes from the 2024-25 academic year and BS2 classes from the 2025-26 academic year.

We will run separate analyses studying whether the type of markup used during TEACH markup affects scoring. Specifically, TEACH training involves four different phases of markup. In order of occurrence, these are:

1. Enumerators markup videos as a group with the master trainer. Scores are an unweighted average across all enumerators
2. Enumerators split into pairs to markup videos. Scores are an unweighted average across both enumerators.
3. Enumerators markup videos individually with group feedback from the master trainer on all videos
4. Enumerators markup videos individually with individual feedback from the master trainer on a randomly selected 20% of the videos.

We will randomize both the schools to be included in each phase and the order in which they are marked.

Obtaining impact estimates

We will obtain experimental impact estimates for lesson quality from TEACH via the following parametric linear model estimated by ordinary least squares.

$$Lesson\ Quality_j = \beta_0 + \beta_1 TFLI_j + Z'_j\tau + X'_i\gamma + \epsilon_{ij} \qquad (1)$$

where *i* indexes students, which are nested within their original schools indexed by j. $TFLI_i$ is the indicator for a school being randomly assigned to receive the TFLI program. Z_{ij} is a vector of

indicators for the stratification cells used in the lottery that assigned schools to study arms. (1) is the specification we will use for all our confirmatory analyses, and we will also use it as our default approach for estimating average treatment effects in any exploratory analyses.

In (1), X_i is a vector of control variables; we will control for an indicator for being male, indicators for each value of age in years (as of the beginning of the academic year)), and the interactions between the two. For students with missing values of any baseline variable, we will replace the missing values with zero and include a separate indicator variable for the original value being zero, along with any appropriate interaction terms.

We will also estimate (2)

$$Y_{ij} = \beta_0 + \beta_1 Lesson\ Quality_j + Z'_j\tau + X'_i\gamma + \epsilon_{ij} \quad (2)$$

using $TFLI_i$ as an instrument for *Lesson Quality*. This will let us estimate how well the program would have worked with full compliance, under the assumption that the effects of the randomized intervention operate only through lesson quality.

We will also analyze if lesson quality is correlated with reading ability within study arms using (3). We will run (3) separately in the treatment and control arms.

$$Y_{ij} = \beta_0 + \beta_1 Lesson\ Quality_j + X'_i\gamma + \epsilon_{ij} \quad (3)$$

Y_{ij} will include all literacy outcomes listed in the main analysis plan.

The TEACH tool is generally used as measures of teacher behavior, but we collect test scores for each student. This gives us two different levels of aggregation at which to estimate treatment effects: student-level scores and teacher-level means. Aggregating to the teacher level discards much of the variation in test scores (and student-level covariates that can explain some of it) but leaving the data at student level implicitly weights treatment effects by class size. To reconcile this, we will report results at both levels of aggregation for these variables. When we report outcomes at the teacher level, we will not control for any covariates, so X'_i will be empty.

For the TEACH markup order analysis section, we will run the following linear regression:

$$Lesson\ Quality_j = \beta_0 + \beta_1 Markup\ Type_j + Z'_j\tau + \epsilon_j \quad (4)$$

$Markup\ Type_j$ is a vector of indicators for which of the 4 types of markup was used for a given lesson observation. Group 1 (group observation) will be the reference group. Z'_j is a vector of indicators for the stratification cell used in the lottery to assign lesson to markup type. We will stratify markup type by treatment status in the main intervention.

Inference

We will conduct inference on (1), (2), and (3) in the same way we analyzed the lesson quality variable we specified in the main analysis plan. Specifically, for (1) and (3) we will conduct inference via randomization inference. We will randomly permute the study arm assignments of each school within the stratification cells used in the original lottery. We will implement this in Stata via the `-ritest-` command. For (2) we will cluster standard errors at the level of the stratification cell.

For (4) we will use HC1 standard errors. Because the data is aggregated at the school level, this is equivalent to clustering at the school level. This randomization was stratified by treatment in the main study. Because the stratification cells are large, it is not necessary to cluster at the stratification cell level. This clustering scheme was validated using simulations.

Null Hypotheses

We will consider one null hypothesis for each outcome we consider.

$$H_0: \beta_1 = 0$$

Outcome

TEACH Primary is organized hierarchically: 28 behaviors nested within 9 elements, which are nested within 3 areas, alongside a separate Time on Learning component.

Each behavior is coded on a three-point scale based on the evidence observed during the lesson: low (1), medium (2), or high (3). Using the behaviors in each element, enumerators then mark each of the 9 elements onto a 1–5 scale: behaviors in the low range correspond to an element score of 1–2, medium to 3, and high to 4–5, with the observer assigning the final value according to the overall quality of the element. The Time on Learning component is coded separately: whether the teacher provides a learning activity to most students (1/0), and the share of students on task, recorded as low (1), medium (2), or high (3). When we aggregate the time on learning area, we will take an evenly weighted average of its two components. We rescale each behavior and element to a 0–1 range by dividing by its maximum possible value (3 for behaviors, 5 for elements). Areas are the evenly weighted average of their rescaled element scores and so are already on a 0–1 scale.

Each element and behavior is scored separately in between 1 and 3 15-minute observation segments. We will average the element and behavior scores across the segments to construct final scores. Time on learning is calculated at each of three snapshots across the observation. A snapshot is a 15-minute segment of the full lesson. We will take the average of the three as the final score. If only one segment (or fewer than three snapshots) is available, we will average over the existing units. We will split videos into segments according to the following rule.

Lesson Video Length	Number of Snapshots
<20	1
20-40	2
>40	3

For the group and pair phases of markup, we will get multiple scores per component: one from each enumerator watching the video. We will take an unweighted average across each enumerator to construct the component score for the videos.

We will use this data in several ways. First, we will report treatment effects separately for each of the nine elements. Second, we will aggregate to area level and report treatment effects for Time on learning, Classroom Culture, Instruction, and Socioemotional Skills separately; we will construct each area to be the evenly weighted average of its element scores. Finally, we will report an overall quality score, which will be the evenly weighted average of all nine element scores and the time on learning score.

The nine elements and their constituent behaviors are listed below.

0. Time on Learning

- Teacher provides learning activity to most students (1/0)
- Students are on task (1-3)

A. Classroom Culture

1. Supportive learning environment

- The teacher treats all students respectfully (1-5)
- The teacher uses positive language with students (1-5)
- The teacher responds to students' needs (1-5)
- The teacher does not exhibit bias and challenges stereotypes in the classroom (1-5)

2. Positive Behavioral Expectations

- The teacher sets clear behavioral expectations for classroom activities (1-5)
- The teacher acknowledges positive student behavior (1-5)
- The teacher redirects misbehavior and focuses on the expected behavior, rather than the undesired behavior (1-5)

B. Instruction

3. Lesson Facilitation

- The teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives (1-5)
- The teacher explains content using multiple forms of representation (1-5)
- The teacher makes connections in the lesson that relate to other content knowledge or students' daily lives (1-5)
- The teacher models by enacting or thinking aloud (1-5)

4. Checks for understanding

- The teacher uses questions, prompts or other strategies to determine students' level of understanding (1-5)
- The teacher monitors most students during independent/group work (1-5)
- The teacher adjusts teaching to the level of students (1-5)

5. Feedback (1-5)

- The teacher provides specific comments or prompts that help clarify students' misunderstandings (1-5)
- The teacher provides specific comments or prompts that help identify students' success (1-5)

6. Critical thinking

- The teacher asks open-ended questions (1-5)
- The teacher provides thinking tasks (1-5)
- The students ask open-ended questions or perform thinking tasks (1-5)

C. Socioemotional skills

7. Autonomy

- The teacher provides students with choices (1-5)
- The teacher provides students with opportunities to take on roles in the classroom (1-5)
- The students volunteer to participate in the classroom (1-5)

8. Perseverance

- The teacher acknowledges the students' efforts (1-5)
- The teacher has a positive attitude towards students' challenges (1-5)
- The teacher encourages goal setting (1-5)

9. Social & Collaborative skills

- The teacher promotes students' collaboration through peer interaction (1-5)
- The teacher promotes students' interpersonal skills (1-5)
- Students collaborate with one another through peer interactions (1-5)