Pre-analysis Plan:
Hybrid Digital Learning Program

Introduction

This project proposes a randomized evaluation of a "hybrid" digital learning program targeted at children in Grades 5-8 in Indian villages. With the COVID-19 pandemic widening the pre-existing learning gap, we hypothesize that programs with appropriate high-quality digital content made available on easily accessible devices and supported by appropriate social support structures can improve learning outcomes and related skills when implemented directly in communities with groups of children. Therefore, this program provides a potential solution to making remote learning more accessible and motivating both during school closures and when schools are open. The evaluation primarily aims to measure the impact of the program on Hindi, English and Math learning outcomes. The non-profit Pratham has developed and piloted the key elements of this intervention and will implement the program to be evaluated in 276 villages in India.

Research Strategy

For this project, data collection has already been carried out, but the analysis has not been conducted. We provide relevant information on the intervention, experimental design and fieldwork below.

Intervention. Pratham, the implementation partner, has been developing a 'Hybrid' program targeted at children in grades 5-8 where high-quality digital content relating to Hindi, English, Math, is made available on easily accessible devices and provided to groups of children in communities and supported by appropriate social support structures.
For this study we test the relative effectiveness of two versions of this program, one 'Full Hybrid' (T1) version and one 'Hybrid' (T2) version. In both interventions, groups of five children receive a tablet that is used during after school working sessions and that features digital content on Hindi, English and Maths. Moreover, a Cluster Resource Leader (hereafter CRL), is responsible for conducting diagnostic tests, uploading content on the tablets, replacing tablets which are not in working condition. CRLs are usually in charge of 5-6 villages, visit each village once a week and brief the students on the next steps for the daily activities for the week. Out of these 5 children, one is assigned the position of group leader and this position changes every 1-2 months.
In the 'Full Hybrid Model' an additional coach is assigned to 4-5 groups (one village) and is a personnel native to the village. The coach is responsible for conducting meetings and activities daily within the study group in their village. In this treatment group, CRLs brief the coaches on the next steps during their weekly visits.

Experimental design. The study was based in the Dausa district in 2 blocks, Mahwa and Bandikui in Rajasthan. These blocks are geographically close to each other, but culturally very different. The 276 study villages were randomized into the following three groups:

Treatment 1		**Treatment 2**		**Control**	
Full Hybrid Model		Hybrid Model		No Intervention	
92 Villages		92 Villages		92 Villages	
Bandikui	Mahwa	Bandikui	Mahwa	Bandikui	Mahwa
52	40	51	41	52	40

Randomization was done at the village and grade level, and stratified by block. Stratification was important to ensure balance across blocks in all treatment groups, given their cultural differences.

Data Collection Timeline. Three phases of data collection were carried out between February 2021 and March 2023. Specifically:

- Baseline was conducted from Feb 12th 2021 to April 16th 2021 and it was interrupted because of the second Covid-19 wave in India. At this stage, only 151 villages out of 155 were covered in the Bandikui block and no data collection was carried out in Mahwa block, yielding no baseline data for the 121 villages in this block.
- Midline was conducted from May 25th 2022 to August 7th 2022.
- Endline was conducted from January 10th 2023 to March 6th 2023.

All data is primary data, collected by surveyors directly from children and their guardians. A typical survey would last about 50 min and children were asked to provide some basic personal information (age, grade, type of school..) before going through the Hindi, English and Maths tests. These tests were carried out using an "illustration sheet" and supervised by the surveyor who would enter the child's responses in the tablet.

If accessible, Pratham's monitoring data might also be used in our analysis.

Outcomes of Interest. As pre-specified in the AEA Registry (RCT ID: AEARCTR-0008330) we use the following outcomes of interest:

Primary outcomes: test scores in Hindi, English and Maths (note that the Science section was replaced by the Hindi section in all data collection). These test scores are measured at a granular level by the calculates b8, b10, b17 (Hindi), c4 to c43 (English) and e1 to e20 (Maths) in the data, which will be aggregated in a pre-specified way to construct the final test scores used for our analysis. Note that between baseline, midline and endline the questions are not identical but are meant to measure the same concept. Merging the data will therefore be done after constructing the relevant composite variables.

Secondary outcomes: because of Covid, none of the secondary outcomes pre-specified in the AEA registry were measured. The analysis will therefore focus on the primary outcomes and explore potential heterogeneous effects.

Empirical Analysis

Once data has been fully cleaned, we anticipate the empirical analysis to unfold in the following way.

Balance checks. To ensure that randomization was effective in creating statistically comparable groups, we will carry out balance checks on our descriptive variables: age, grade, type of school, block, baseline test scores. Balance will also be tested on survey attrition. Note that baseline scores are not available for villages in the Mahwa block as explained in the *Data Collection Timeline* section, hence balance checks will be carried out both for the full sample and at the block level. We aim to perform these balance tests both:

- as standard comparisons of means;
- as regressions to include stratification.

As specified below, we will control for most of these variables in our main regression. This allows us to increase the precision of our estimates, and will appear necessary if important imbalances arise across treatment groups. To correct for potential imbalances, we might alternatively use entropy weights (Hainmueller 2012) to match groups on all observables we have.
Finally, if samples change for certain specifications due to missing data, data quality issues or other reasons, we will perform balance checks on those samples relevant to our analysis and robustness checks.

Main analysis. Our main analysis seeks to explore the causal impact of each treatment arm on Hindi, English and Maths learning outcomes. The main regression will be

$$y_i = \alpha + \beta T_1 + \gamma T_2 + \delta X_i + \eta FE_b$$

where:

- y_i are child-level composite test scores in either Hindi, English or Maths depending on the specification. Broadly speaking, these composite variables will be constructed in the following way (see table below for additional details):
    - Hindi: y_i is a scale ranging from 1-5 characterizing whether the child is at beginner (1), letter (2), word (3), paragraph (4) or story (5) level.
    - English: y_i is a general English score constructed from sub-indexes of vocabulary, listening; reading etc.. To explore the program's impact on specific sub-skills, these sub-indexes will also be used as dependent variables.

    - Maths: y_i will be a score out of the total number of questions (20 for endline). For compatibility across baseline, midline and endline the math score might be rescaled to same-length intervals.
- T_1 and T_2 are indicator variables for the two treatment groups. In some specifications these two might be pooled to gain power.
- X_i are child-level controls such as age, grade, type of school and baseline scores.
- FE_b are block-level fixed effects. If available, we could use Pratham's monitoring data to construct coach-level fixed effects.
- Standard errors will be clustered at the level of randomization i.e. village & grade.

The table below summarizes how composite variables will be constructed in practice (for endline data).

<table>
<tr><th></th><th>Survey questions used for aggregation</th><th>What is measured ?</th><th>Composite variable & Aggregation rule</th></tr>
<tr><td rowspan="9">Hindi</td><td>b5_able_to_read</td><td rowspan="3">Was the child able to read the story (3 attempts). Coded as 1 if successful, 0 otherwise.</td><td rowspan="9">The composite variable hindi_level will be constructed as follows:
1 (beginner): if the child is unable to read story, paragraph, word and letters.
2 (letter): if the child is able to read letters but no words.
3 (word): if the child is able to read words but no paragraph.
4 (paragraph): if the child is able to read paragraph but no story.
5 (story): if the child is able to read everything.</td></tr>
<tr><td>b13_able_to_read</td></tr>
<tr><td>b19_able_to_read</td></tr>
<tr><td>b4_able_read_para</td><td rowspan="3">Was the child able to read the paragraph (3 attempts). Coded as 1 if successful, 0 otherwise.</td></tr>
<tr><td>b12_able_read_para</td></tr>
<tr><td>b18_able_read_para</td></tr>
<tr><td>hindi_word_1</td><td rowspan="2">Was the child able to read the word (2 attempts). Coded as 1 if successful, 0 otherwise.</td></tr>
<tr><td>hindi_word_2</td></tr>
<tr><td>b9_how_many_letter_read</td><td>Was the child able to read 4 letters ?</td></tr>
<tr><td rowspan="6">English</td><td>c1_fruits_a_1 to c1_fruits_a_6</td><td>Did the child name fruits 1 to 6 ?</td><td rowspan="3">Composite variable vocabulary A: total score out of the 18 identification questions.</td></tr>
<tr><td>c2_vegetables_a_1 to c2_vegetables_a_6</td><td>Did the child name vegetables 1 to 6 ?</td></tr>
<tr><td>c3_animals_a_1 to c3_animals_a_6</td><td>Did the child name animals 1 to 6 ?</td></tr>
<tr><td>c4_who_this</td><td rowspan="3">Did the child identify the profession ? This is re-coded as 0 (wrong answer / no answer) or 1 (correct answer).</td><td rowspan="3">Composite variable for vocabulary B: total score out of the 3 professions.</td></tr>
<tr><td>c5_who_this</td></tr>
<tr><td>c6_who_this</td></tr>
</table>

	c7_he_doing c8_she_doing c9_he_doing	Did the child identify the activity ? This is re-coded as 0 (wrong answer / no answer) or 1 (correct answer).	Composite variable for **vocabulary C:** total score out of the 3 activities.
	c10_word c11_word c12_word	Did the child translate the word correctly into English ? This is re-coded as 0 (wrong answer / no answer) or 1 (correct answer)	Composite variable for **translation:** total score out of the 3 translations.
	c15_bro_sis c17_class c19_birthday c13_village_name	Did the child answer the four questions asked in English ? This is re-coded as 0 (wrong or no answer), 0.5 (partially correct) and 1 (correct).	Composite variable for **speaking:** total score out of the 4 questions.
	c24_read_first_word c25_read_second_word c26_read_third_word c27_read_fourth_word c28_read_fifth_word	Was the child able to read words 1-5 ? This is re-coded as 0 (no answer / no) and 1 (yes).	Composite variable for **reading words:** total score out of the 5 questions.
	c21_school c22_kamla c23_jaipur	Was the child able to tell the meaning in English of the three sentences heard ? This is re-coded 0 (wrong or no answer), 0.5 (partial meaning) and 1 (complete meaning).	Composite variable for **telling meaning in English A:** total score out of the 3 questions.
	c29_read_1st_sent c31_read_2nd_sent c33_read_3rd_sent	Was the child able to read sentences ? This is re-coded as 0 (wrong or no answer), 0.5 (partially fluent) and 1 (fluent).	Composite variable for **reading sentences:** total score out of the 3 questions.
	c30_able_tell_meang c32_able_tell_meang c34_able_tell_meang	Was the child able to tell the meaning of the sentences ? This is re-coded as 0 (wrong or no answer), 0.5 (partial meaning) and 1 (complete meaning).	Composite variable for **telling meaning in English B:** total score out of the 3 questions.
	c35_read_paragrah	Was the child able to read the paragraph ? This is re-coded as 0 (wrong or no answer), 0.5 (partially fluent) and 1 (fluent).	Composite variable for **reading paragraph & meaning**: total score out of the 4 questions.

<table>
<tr><td rowspan="7"></td><td>c36_new_dress</td><td rowspan="3">Was the child able to answer follow-up questions ? This is re-coded as 0 (wrong or no answer), 0.5 (partially correct) and 1 (correct).</td><td rowspan="3"></td></tr>
<tr><td>c37_radha</td></tr>
<tr><td>c38_likes_dress</td></tr>
<tr><td>c39_read</td><td>Was the child able to read the pamphlet ? This is re-coded as 0 (wrong or no answer), 0.5 (partially fluent) and 1 (fluent).</td><td rowspan="4">Composite variable for reading pamphlet & meaning: total score out of the 4 questions.</td></tr>
<tr><td>c41_used_for</td><td rowspan="3">Was the child able to answer follow-up questions ? This is re-coded as 0 (wrong or no answer), 0.5 (partially correct) and 1 (correct).</td></tr>
<tr><td>c42_good</td></tr>
<tr><td>c43_used</td></tr>
<tr><td colspan="4">Seven sub-indexes will be created using simple aggregation, in order to explore the program’s impact on specific skills:
- Vocabulary, Translation, Speaking, Reading & meaning words, Reading & meaning sentences, Reading & meaning pamphlet, Reading & meaning paragraph.

A final index will be constructed by aggregating those sub-indexes in the following way:
- Vocabulary + Translation using simple aggregation to create eng1
- Speaking to create eng2
- Reading & meaning words + sentences using simple aggregation to create eng3
- Reading & meaning pamphlet + paragraph using simple aggregation to create eng4

Each eng1-4 sub-index will be standardized and the final English level index will be constructed using a weighted average of those standardized sub-indexes, using equal weights ¼.

As a robustness check, and to have a more data-driven procedure for weight selection, an alternative English level index will be created in the following way:
- Standardize the 7 sub-indexes
- Use PCA to compute weights on each sub-index
- Take a weighted average to construct the final index.</td></tr>
<tr><td rowspan="8">Maths</td><td>e1_fraction_b</td><td rowspan="8">Each variable measures whether the child answered correctly to the corresponding question. The questions cover topics about fractions, angles, geometry, calculations, logic and solving short problems. Note that these variables have been re-coded to 0 (wrong answer) and 1 (correct answer).</td><td rowspan="8">The composite variable math_score will be constructed by summing all points obtained to the individual questions. It will be a score out of 20.</td></tr>
<tr><td>e2_angle</td></tr>
<tr><td>e3_area</td></tr>
<tr><td>e4_solve</td></tr>
<tr><td>e5_solve</td></tr>
<tr><td>e6_fraction_orange</td></tr>
<tr><td>e7_solve</td></tr>
<tr><td>e8_triangle</td></tr>
</table>

	e9_mahesh		
	e10_cotton		
	e11_schoolbus		
	e12_farmperimeter		
	e13_bluehouse		
	e14_bluehouse		
	e15_triangle		
	e16_number_line		
	e17_sarita		
	e18_hyp		
	e19_paper_b		
	e20_map_b		

Again, since baseline data is not available for the Mahwa block, we anticipate to run this regression on the full sample (dropping baseline scores) as well as at the block level.

Attrition and missing data. If we find any differential attrition, we will conduct robustness checks to confirm our results.
If we have missing values among control variables, we aim to drop these observations in our analysis. If our results seem to be underpowered because of missing values we will 1) run our analysis without controls to maximize sample size or 2) input missing data using standard matching methods.
If we have missing values for test scores, these will affect the construction of our composite variables. Hence we anticipate the following procedure:

- remove observations with missing test data and run the analysis
- as a robustness check, or if results are underpowered due to removing those observations, include them wherever possible. This could be done by creating a "thinner" version of the composite variables keeping only questions with a small share of missing data, running regressions on certain sub-indexes (e.g. for the English section) or even at the question level.

Outliers. We anticipate two types of outliers in our data:

- outliers with extreme values compared to the rest of the distribution, yet these values are plausible. In that case it is important to keep them in our main analysis as they carry some representativeness of how a real-life sample could look like.

- outliers with extreme values compared to the rest of the distribution, yet these values are impossible or extremely implausible. In that case the main analysis should drop these observations. As a robustness check (or if dropping leads to a notable loss in power), these variables will be winsorized or replaced by the modal value across a group of similar observations (except for outcome variables which are never inputted).

Given our data, age (*calc_age*) and grade (*calc_grade*) are the only variables for which we might encounter outliers. This motivates the following rule:

- An "implausible" outlier is a child reporting an age 4 or more years apart from the expected age in the reported grade. For example, a child in grade 8 is expected to be at most 14. Hence, a child reporting to be in grade 8 with a reported age of 18 or more is considered an "implausible" outlier. In that case we replace their age and grade variable with missing values.

Note that scores for any Hindi, English or Maths question can range from 0 to maximum points. Any value outside this range will be replaced by missing, and any value within is considered possible, so that outliers are not an issue here.

Heterogeneity analysis. Our secondary analysis will focus on exploring heterogeneous treatment effects by:

- Gender (measured by p14_gender)
- Age (measured by calc_age)
- Type of school (measured by p18_school_type: 1. Government, 2. Private, 998. Others)
- Grade (measured by calc_grade)
- Block (measured by calc_block)

Additionally, the main regression will be interacted with (data collection) "phase" to measure the program's differential impact at midline and endline.

Note that given the Covid-19 context, many children have stopped going to school for several months and were not necessarily directed to the appropriate "grade" upon return. During data collection, the grade variable thus turned out to be very hard to reconcile with other data, despite having several questions aimed at identifying the child's grade. We therefore anticipate the analysis by grade to be noisy and interpretation will need to be carefully made.

Similarly, the block-level results will be interpreted carefully given the absence of a baseline for one of the two blocks.

Cost-effectiveness analysis. Finally, a cost-effectiveness analysis will be performed using Pratham's data on implementation costs. Costs related to the evaluation will not be included as part of this analysis.

References

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political analysis, 20(1), 25-46.