***FOURTH ENDLINE ANALYSIS: Every Child Counts***

General Notes:

1. This document outlines our main tests regarding causal effects of the treatment at the fourth endline. We will also do complementary analysis to both describe and examine the channels/mechanisms through which children learn.
2. Practice trials are never included in any of these analyses.
3. Responses of “I don’t know”, “refuse to answer” or trials skipped due to actions of the child (such as crying) are coded as incorrect; trials are not scored if the child does not finish the test for reasons beyond their control (such as a computer failure).
4. Percentage of correct responses is calculated by dividing the number of correct trials by the number of trials given (i.e., correct/(correct + incorrect + I don't know + refuse to answer)). Trials that are not given due to circumstances beyond the child’s control do not enter into this calculation.
5. Z-scores are calculated by standardizing to the contemporaneous control values.
6. Throughout we use the following regression frameworks:

(1)

(2)

Where represents the endline value of an outcome for child *i* in school *j*,is a indicator variable for whether school *j* was treated with the math games intervention, is an indicator variable for whether school *j* was treated with the social games intervention, is age in months of child *i* in school *j,*  is gender of child *i* in school *j*, are a set of baseline test scores for child *i* in school *j* whose functional form has been selected using the post-double lasso procedure of Belloni, Chernozhukov and Hansen (2014) from a set of principle components, geometry and number Z-Scores, and symbolic and non-symbolic Z-scores, and the individual tests*,* and is an indicator for whether the child was surveyed during the pre-COVID wave. Standard errors will be clustered at the school level.

Overall Outcomes and Hypotheses:

Main Outcome 1: Z-Score of percentage correct across all math tests (Geometric Intruder, Panamath, Numerical Reasoning, Reading Maps, Fluency, Geometric Reasoning, Symbolic Approximate Arithmetic, ASER Math)

* Hypothesis:
  + The math treatment will generate positive effects and the social treatment will generate smaller effects.
* Test:
  + Regress outcomes using (1) and (2)

Main Outcome 2: Z-Score of percentage correct across all *non-symbolic* math tests (Geometric Intruder, Panamath, Reading Maps)

* Hypothesis:
  + The math treatment will generate positive effects and the social treatment will generate smaller effects.
* Test:
  + Regress outcomes using (1) and (2)

Main Outcome 3: Z-Score of percentage correct across all *symbolic* math tests (Numerical Reasoning, Fluency, Geometric Reasoning, Symbolic Approximate Arithmetic, ASER MATH)

* Hypothesis:
  + The math treatment will generate positive effects and the social treatment will generate smaller effects.
* Test:
  + Regress outcomes using (1) and (2)

Main Outcome 4: Z-Score of percentage correct across all social tests (Gaze Direction, Identify Emotion, Emotion Intruder)

* Hypothesis:
  + The social treatment will generate positive effects and the math treatment will generate smaller effects.
* Test:
  + Regress outcomes using (1) and (2)

Main Outcome 5: Z-score of score out of 4 on English ASER. (For all the ASER’s we will add an additional non-standard level of “partially knows” the lowest level, whether that’s single letter or single digit)

* Hypothesis:
  + If the math or social treatments generate positive effects on the symbolic math outcome (Main Outcome 3), then the corresponding treatment will generate smaller positive effects on this outcome due to either an increased interest in schooling or a greater familiarity with symbols in general. Otherwise, there will be no effect of either the math or the social treatment.
* Test:
  + Regress outcomes using (1) and (2)

Test Specific Outcomes and Hypotheses:

Geometric Intruder: There are 4 practice trials and 10 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: This is the most direct test of the math treatment. The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

Panamath: There are 4 practice trials and 12 test trials. Practice trials may be repeated either before each block (if children get them wrong) or in the middle of a block (if children perseverate on one response side).

* Main outcome: Percentage of correct responses
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

Numeric Reasoning: There are 4 practice trials and 16 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

Geometric Reasoning: There are 2 practice trials and 8 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

Gaze Direction: There are 0 practice trial and 4 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: This is the most direct test of the social treatment. The social intervention will generate positive effects and the math intervention will generate smaller effects
* Test: Regress outcomes using (1) and (2).

Identify Emotion: There are 0 practice trials and 5 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: The social intervention will generate positive effects and the math intervention will generate smaller effects
* Test: Regress outcomes using (1) and (2).

Emotion Intruder: There are 2 practice trials and 5 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: The social intervention will generate positive effects and the math intervention will generate smaller effects
* Test: Regress outcomes using (1) and (2).

Reading Maps: There are 4 practice trials and 8 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

Fluency: There are 0 practice trials and 1 test trial.

* Main outcome: Percentage of correct responses
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

Symbolic Approximation Arithmetic: There are 0 practice trials and 5 test trials.

* Main outcome: Percentage of correct responses
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

ASER Math: There are between 2 and 3 trials depending on performance and no practice.

* Main outcome: Score out of 4. Subjects will get zero for not knowing any single digit numbers, half a point for knowing some but not enough single digit numbers to pass that level, one for knowing single digit numbers, two for knowing double digit numbers, three for knowing subtraction, four for knowing division.
* Hypothesis: The math intervention will generate positive effects and the social intervention will generate smaller effects.
* Test: Regress outcomes using (1) and (2).

ASER English: There are between 2 and 3 trials depending on performance and no practice.

* Main outcome: Score out of 4 calculated for English. Score out of 4. Subjects will get zero for not knowing any letters, half a point for knowing some but not enough letters to pass that level, one for knowing letters, two for knowing words, three for knowing paragraph, four for knowing story.
* Hypothesis: If the math or social treatments generate positive effects on the symbolic math outcome (Main Outcome 3), then the corresponding treatment will generate smaller positive effects on this outcome due to either an increased interest in schooling or a greater familiarity with symbols in general. Otherwise, there will be no effect of either the math or the social treatment.
* Test: Regress outcomes using (1) and (2)

Additional Outcomes and Hypotheses:

Motivation

* Main Outcomes: Level of difficulty chosen for additional math questions and whether math was chosen over English or Executive Function for the additional questions.
* Other Outcomes of Interest: Stated levels of motivation for mathematics and levels of difficulty chosen for the other domains.
* Hypothesis: The math intervention will cause subjects to choose a higher level of math difficulty and increase the likelihood of choosing math.
* Tests:
  + Regress the outcomes using (1) and (2).
  + In order to look at whether this is mediated by performance on the current test, we will regress (2) while interacting the treatment indicators with the all mathematics Z-score. This specification suffers from potentially bad controls if we find treatment increases math performance, but we still think this is an interesting secondary analysis.

Additional Correlations: In order to better understand the results in context, we will also estimate intra- and intertemporal correlations between Z-score indexes of number and geometry tests.

Examining Heterogeneity in Treatment Effects:

* Main Outcome: Z-Score of percentage of correct responses across all math tests examined by baseline executive function performance, baseline math test performance, and age.
* Hypotheses:
  + Of students in the math intervention, those who performed better on the baseline executive function test will have larger treatment effects.
  + The relative sizes of treatment effects based on baseline math assessment scores and age are uncertain, but policy relevant.
  + Students with more exposure to school will have larger treatment effects on both outcome measures and the difference will be larger on the symbolic math outcome.
* Tests:
  + Regress the main outcome on an interaction of the treatment dummies with the percentage correct on blocks 2, 3, and 4 from the baseline executive function test and main effects. Any other analysis of executive function will also use this percent-correct measure rather than the previous response-switching and task-switching measures due to their low test-reliability discovered after EL2.
  + Separate students at terciles of the Z-score of percentage of correct responses across all math tests at baseline (Panamath, Point to Number, Point to Shape and Geometric Intruder) and regress the main outcome using (1) and (2).
  + Regress the main outcome on an interaction of the treatment dummies with age and main effects.