

## A Supplementary Tables and Figures

Table A.1: Evidence on Pedagogy and Declines in Performance from the TIMSS Exam

	Time Spent on Indep. Practice		Item Correct	
	(1)	(2)	(3)	(3)
Log GDP	0.0105*** (0.00230)			0.0704*** (0.00199)
Question Location		-0.0393*** (0.00376)		-0.133*** (0.0121)
Time Spent on Indep. Practice		0.0885*** (0.0204)		0.0695*** (0.0202)
Question Location x Time Spent on Indep. Practice		0.0317*** (0.00682)		0.0278*** (0.00700)
Question Location x Log GDP				0.00954*** (0.00113)
Constant	0.422*** (0.0224)	0.513*** (0.0107)		-0.175*** (0.0218)
Observations	7476337	8217081		7476337

*Notes:* This table uses TIMSS exam data and GDP data collected from the World Bank (2015) to show both the differences in pedagogy by income and differences in rate of decline of performance by economic advantage and pedagogy. We conduct this analysis within the global sample, where there is greater power to examine these correlations due to the larger sample size.

- Observations are at the student-question level. Question item order is block randomized. The dependent variable in Col. (1) is the fraction of study time students spend working independently, and in Cols. (2)-(3) is whether a given question item is answered correctly.
- “Log GDP” is log of GDP/capita (PPP) of the student’s country of residence in the year of the exam from World Bank (2015). “Question Location” is a variable from 0 to 1 which captures the location of the question item in the test, where 0 is the beginning and 1 is the end. “Time Spent on Indep. Practice” is the fraction of study time students spend working independently drawn from the teacher survey in TIMSS.
- Col. (1) tests the relationship between country income and time spent on independent practice. Cols. (2)-(3) document the relationship between country income and independent practice time on the rate of decline over the test.
- All regressions contain question fixed effects. Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.2: Baseline Balance

Variable	(1) Control		(2) Cog. Practice (Pooled)		(3) Games Practice		(4) Math Practice		(5)	(6) T-test p-value		(7)	(8)
	N	Mean/SE	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(1)-(4)	(3)-(4)	
<b>Panel A: Student Characteristics</b>													
Grade	537	2.767 (0.063)	1091	2.720 (0.044)	543	2.742 (0.063)	548	2.697 (0.062)	0.535	0.778	0.427	0.609	
School Income Tercile	537	2.255 (0.036)	1091	2.258 (0.025)	543	2.273 (0.035)	548	2.245 (0.035)	0.938	0.727	0.832	0.573	
Income Tercile	537	1.959 (0.024)	1091	1.959 (0.017)	543	1.974 (0.024)	548	1.943 (0.023)	0.992	0.654	0.640	0.357	
Baseline Ability Tercile	537	1.991 (0.028)	1091	1.995 (0.020)	543	1.994 (0.028)	548	1.996 (0.028)	0.891	0.925	0.886	0.962	
Female	530	0.349 (0.021)	1074	0.363 (0.015)	536	0.351 (0.021)	538	0.375 (0.021)	0.581	0.954	0.370	0.400	
<b>Panel B: Student Baseline Scores</b>													
Baseline Listening (mean)	493	0.563 (0.017)	1011	0.551 (0.012)	500	0.546 (0.017)	511	0.556 (0.017)	0.594	0.501	0.799	0.672	
Baseline Math (mean)	485	0.399 (0.010)	981	0.412 (0.007)	486	0.413 (0.010)	495	0.412 (0.009)	0.267	0.341	0.334	0.990	
Baseline Ravens Matricies (mean)	483	0.369 (0.012)	985	0.373 (0.009)	483	0.364 (0.012)	502	0.382 (0.012)	0.791	0.746	0.446	0.279	
Baseline Listening (decline)	482	-0.003 (0.020)	985	-0.019 (0.015)	488	-0.024 (0.021)	497	-0.015 (0.020)	0.510	0.465	0.673	0.749	
Baseline Math (decline)	485	-0.070 (0.017)	981	-0.064 (0.012)	486	-0.073 (0.017)	495	-0.055 (0.017)	0.778	0.894	0.533	0.457	
Baseline Ravens Matricies (decline)	454	-0.075 (0.025)	936	-0.036 (0.022)	454	-0.024 (0.040)	482	-0.046 (0.019)	0.279	0.289	0.369	0.614	

*Notes:* This table presents summary statistics for student baseline covariates by treatment group and tests whether there is any statistically significant difference between experimental arms at baseline.

- Cols. (1), (2), (3), and (4) present the sample size and mean for each covariate by treatment status. Column (2) pools students across both sub-treatments, and Cols. (3) and (4) separate students by their sub-treatment. Cols. (5)-(8) present p-values for the test of equality of means between the indicated groups.

- Panel A includes student characteristics. Student baseline ability—proxied by average baseline math grades—and income were provided by the school. Students' gender was determined based on their name. For 1.6% of the sample, the student's name was gender neutral, so we leave the variable missing. School Income Tercile is based on the school the student attended, with two of the six schools designated in each tercile.

- Panel B presents students' performance in each of the performance decline tests, showing the average score on the test (mean) and the performance at the end of the task minus the beginning (decline). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.3: Baseline Balance: School Performance

	Treatment			Disaggregated Sub-Treatments				
	Control	Cog. Practice	p-value 1 = 2	Math Practice	Games Practice	p-value 1 = 4	p-value 1 = 5	p-value 4 = 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pooled	0.02 (0.039)	-0.01 (0.032)	0.71	-0.10 (0.045)	0.08 (0.044)	0.08	0.12	0.01
Hindi	0.02 (0.092)	-0.07 (0.077)	0.62	-0.17 (0.108)	0.03 (0.109)	0.25	0.63	0.19
English	0.01 (0.091)	0.02 (0.075)	0.84	-0.07 (0.107)	0.11 (0.106)	0.62	0.33	0.23
Math	0.02 (0.091)	-0.03 (0.070)	0.76	-0.08 (0.099)	0.03 (0.098)	0.55	0.82	0.42

*Notes:* This table presents summary statistics for student baseline grades on school administered tests by treatment group and shows whether there is any statistically significant difference between experimental arms at baseline.

- Data is available for only two schools with a total of 95 Control students, 87 Math treatment students, and 86 Games treatment students.
- The dependent variable is the standardized z-score of the student's class grade from the previous year.
- Cols. (1), (2), (4), and (5) present the mean and standard error for each covariate by treatment status. Col. (2) pools students who are in either sub-treatment, and Cols. (4) and (5) separate students by their sub-treatment. Cols. (3) and (6)-(8) present p-values for the test of equality of means between the indicated groups.
- "Pooled" includes all three academic subjects. Hindi, English, and math respectively examine each class subject individually.
- "Cog. Practice" is a dummy for whether the student was in either sub-treatment group. "Math Practice" is a dummy for whether the student was part of the math sub-treatment and "Games Practice" is a dummy for whether the student was part of the games sub-treatment.
- \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.4: Attrition

	Treatment			Disaggregated Sub-Treatments				
	Control	Cog. Practice	p-value 1 = 2	Math Practice	Games Practice	p-value 1 = 4	p-value 1 = 5	p-value 4 = 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A: Declines Tests: Listening, Ravens, and Math</b>								
Pooled	0.9708 (0.006)	0.9777 (0.004)	0.32	0.9778 (0.005)	0.9776 (0.005)	0.39	0.41	0.98
Math	0.9615 (0.007)	0.9744 (0.004)	0.09	0.9752 (0.006)	0.9736 (0.006)	0.13	0.18	0.85
Listening	0.9668 (0.007)	0.9692 (0.004)	0.76	0.9687 (0.006)	0.9697 (0.006)	0.84	0.75	0.91
Ravens	0.9602 (0.007)	0.9698 (0.004)	0.23	0.9687 (0.006)	0.9710 (0.006)	0.37	0.25	0.79
<b>Panel B: Psychological Measures of Attention</b>								
Pooled	0.0000 (0.000)	0.0000 (0.000)	1.00	0.0000 (0.000)	0.0000 (0.000)	1.00	1.00	1.00
Symbol Matching	0.9969 (0.002)	0.9962 (0.002)	0.80	0.9939 (0.003)	0.9985 (0.002)	0.43	0.56	0.18
SART	0.9089 (0.013)	0.9377 (0.008)	0.05	0.9289 (0.012)	0.9469 (0.010)	0.26	0.02	0.25
<b>Panel C: School Grades</b>								
Pooled	0.8957 (0.011)	0.8955 (0.008)	0.99	0.8980 (0.011)	0.8930 (0.011)	0.88	0.87	0.75
Math	0.8957 (0.011)	0.8955 (0.008)	0.99	0.8980 (0.011)	0.8930 (0.011)	0.88	0.87	0.75
Hindi	0.8957 (0.011)	0.8955 (0.008)	0.99	0.8980 (0.011)	0.8930 (0.011)	0.88	0.87	0.75
English	0.8957 (0.011)	0.8955 (0.008)	0.99	0.8980 (0.011)	0.8930 (0.011)	0.88	0.87	0.75
<b>Panel D: School Enrollment in Following Year</b>								
Pooled	0.8867 (0.012)	0.8965 (0.008)	0.25	0.9060 (0.011)	0.8868 (0.012)	0.11	0.79	0.11

*Notes:* This table presents the extent of attrition for each student outcome and tests whether attrition is differential by treatment status.

- The outcome is whether we observe at least one (non-baseline) test per student.
- Cols. (1), (2), (4), and (5) present the percent of students for whom we have the respective exam. Cols. (3) and (6)-(8) present a p-value for the test of whether attrition is differential by treatment.
- Panel A presents the results for the listening, Ravens Matrices, and math tests. Panel B is for the psychological measures of attention (SART and Symbol Matching). Panel C provides data for the school-administered end-of-term grades. Panel D examines re-enrollment in the year following the study as proxied by observing at least one test in the follow-up round of testing the next year, conditional on having been observed at midline or endline during the experimental year. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.5: Treatment Effects on School Performance Results Disaggregated by Round

	Dependent Variable: Z-score			
	Test Subject			
	All	Non-Math	All	All
	(1)	(2)	(3)	(4)
Cog. Practice x End of Year	0.0940*** (0.0357)	0.1018*** (0.0388)		
Cog. Practice x Mid-Year	0.0737** (0.0370)	0.0682* (0.0398)		
Math Practice x End of Year			0.0897** (0.0411)	
Math Practice x Mid-Year			0.0763* (0.0428)	
Games Practice x End of Year				0.0984** (0.0408)
Games Practice x Mid-Year				0.0711* (0.0424)
F-test p-value Diff. of 2 coefficients = 0	0.4208	0.1974	0.6505	0.3373
Observations	11320	7539	11320	11320

*Notes:* This table reports treatment effects on students' regular school performance, disaggregated by mid-year and end-of-year grades, in the three core subjects offered by all schools in the study (Hindi, English, and math).

- Observations are at the student-subject-semester level. The dependent variable is the standardized z-score of the student's class grade.
- "Cognitive Practice" denotes receiving either treatment. "Math Practice" and "Games Practice" denote the Math or Games sub-treatments, respectively.
- Cols. (1), (3) and (4) include all three subjects. Col. (2) restricts to non-math subjects (English and Hindi).
- All regressions include class section fixed effects and a linear control for baseline school performance. Standard errors clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.6: Heterogeneous Treatment Effects on School Performance and Performance Declines

**Panel A: School Tests**

<i>Covariate:</i>	<b>Dependent Variable: Z-score of Student's Grades</b>			
	(1) Grade	(2) Female	(3) Baseline Mean	(4) Baseline Decline
Cog. Practice	0.1044 (0.0802)	0.0731* (0.0438)	0.1052 (0.0899)	0.0943*** (0.0343)
Cog. Practice x Covariate	-0.0068 (0.0244)	0.0303 (0.0688)	-0.0780 (0.1614)	-0.2182 (0.1665)
p-value: Cog. Practice + Cog. Practice x Covariate = 0	0.0981	0.0520	0.7421	0.4524
Observations	11320	11320	11320	11320

**Panel B: Decline on Listening, Math and Ravens Tests**

<i>Covariate:</i>	<b>Dependent Variable: 1[Question Correct]</b>			
	(1) Grade	(2) Female	(3) Baseline Mean	(4) Baseline Decline
Cog. Practice x Predicted Decline	0.1440** (0.0719)	0.0884** (0.0372)	0.1183 (0.0895)	0.0764*** (0.0286)
Cog. Practice x Predicted Decline x Covariate	-0.0179 (0.0197)	0.0048 (0.0577)	-0.0672 (0.1585)	0.0673 (0.1516)
Dep. Var. Mean	0.47	0.47	0.47	0.47
p-value: Cog. Practice x Pred. Decline + Cog. Practice x Pred. Decline x Covariate = 0	0.0200	0.0321	0.5117	0.3426
Observations	329349	329349	329349	329349

*Notes:* The table shows whether there was a heterogeneous treatment effect of the program by student covariate on students' school grades (Panel A) or performance decline (Panel B).

- Panel A regressions are based on the specification in Table I, Col. (1); Panel B regressions are based on the specification Table II, Col. (2). Each regression estimates heterogeneous treatment effects by the student covariate in the given column.

- "Covariate" varies by column. In Column (1), it is grade, which ranges from 1-5. In Column (2), it is a binary indicator for whether the student is female. In Column (3), it is student's baseline average percent of questions correct on the listening, math, and Ravens tests, and in Column (4), it is the average difference between student's performance in the first quintile of the test versus the last quintile on the baseline tests.

- The other variable definitions, controls, and specifications are identical to that of Table I for Panel A and Table II for Panel B. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.7: Test Completion Rates

	Math (1)	Listening (2)	Ravens (3)
% attempted	0.794	1.000	0.988
% skipped	0.140	0.000	0.004
% of students completing last question item	0.708	1.000	0.984
Avg last question completed location	0.920	0.968	0.981

*Notes:* This table presents information about how much of each test students completed for all non-baseline tests.

- Each row lists a statistic which captures features of student test completion. “% attempted” is the percentage of individual question items students provided an answer to. “% skipped” is the percent of questions in which students left a question blank but answered at least one subsequent question. “% of students completing last question item” captures the percent of students who provided an answer on the last question of the exam, proxying for “finishing” the exam. “Avg. last question completed location” captures the average location of the last question item a student completed on the test as a percent of the total test.
- Each column shows the given statistic for the respective test: math, listening and Ravens. The listening and Ravens tests are multiple-choice tests, and the math exam is free response.

Table A.8: Test Completion Balance

	Treatment			Disaggregated Sub-Treatments				
	Control	Cog. Practice	p-value 1 = 2	Math Practice	Games Practice	p-value 1 = 4	p-value 1 = 5	p-value 4 = 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pooled	96.22 (0.161)	96.24 (0.112)	0.93	96.24 (0.158)	96.24 (0.158)	0.95	0.93	0.99
Math	91.76 (0.371)	92.01 (0.253)	0.56	92.23 (0.355)	91.79 (0.361)	0.35	0.95	0.38
Listening	98.97 (0.124)	98.93 (0.095)	0.80	98.99 (0.127)	98.87 (0.142)	0.91	0.58	0.51
Ravens	98.74 (0.178)	98.59 (0.129)	0.51	98.15 (0.217)	99.03 (0.139)	0.04	0.19	0.00

*Notes:* This table tests whether there is balance in test completion by experimental arm.

- Data is at the student level and is from the three tests used to measure student declines: math, listening, and Ravens Matrices.
- The outcome variable is the location of the last question completed, where the test length is normalized to begin at 0 and end at 100.
- Cols. (1), (2), (4), and (5) present the mean location of the last question completed and corresponding standard error. Cols. (3) and (6)-(8) present a p-value for the test of whether the last location reached is balanced by treatment.
- “Cog. Practice” is a dummy for whether the student was in either treatment group. “Math Practice” is a dummy for whether the student was part of the math sub-treatment and “Games Practice” is a dummy for whether the student was part of the games sub-treatment.

Table A.9: Test Length Summary

Test	Length (minutes)
Math	30
Listening	12-15
Ravens	15-20
SART	8
Symbol matching	15

*Notes:* This table reports the length of tests administered by the research team.

Table A.10: Balance in Question Difficulty by Question Location and Treatment

<i>Test Quintile:</i>	<b>Dependent Variable:</b>				
	<b>% of control group which answered question correctly</b>				
	1	2	3	4	5
Math Practice	-0.00571 (0.231)	-0.00493 (0.256)	0.00128 (0.659)	-0.000778 (0.854)	-0.00355 (0.397)
Games Practice	0.00166 (0.734)	0.00303 (0.473)	0.00176 (0.545)	-0.00650 (0.129)	0.00155 (0.726)
Dep. Var. Mean	0.490	0.443	0.456	0.443	0.468
Observations	77148	52698	71712	52698	80266

*Notes:* This table tests whether there is balance in question item difficulty in each quintile of the test by treatment status.

- Data is at the student-question item level and is from the three tests used to measure student declines: math, listening, and Ravens Matrices. The outcome variable is the control group mean performance on the given question item.
- “Math Practice” is a dummy for whether the student was part of the math sub-treatment and “Games Practice” is a dummy for whether the student was part of the games sub-treatment.
- Each column restricts to question items in the given quintile of the test (e.g., column (1) restricts to the first 20% of the test).
- Standard errors are clustered at the student level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.11: Treatment Effects on Performance Declines — Robustness to Alternate Definitions of Predicted Decline

	<b>Dependent Variable: 1[Question Correct]</b>		
<i>Predicted Decline Calculated:</i>	By Quintile (1)	by Quintile-School (2)	by Quintile-School-Test (3)
<b>Panel A: Pooled Treatment Arms</b>			
Cog. Practice x Predicted Decline	0.0675** (0.0273)	0.0898*** (0.0281)	0.0643** (0.0273)
Cog. Practice	-0.0010 (0.0057)	-0.0046 (0.0060)	0.0014 (0.0052)
<b>Panel B: Disaggregated Treatment Arms</b>			
Math Practice x Predicted Decline	0.0678** (0.0309)	0.0947*** (0.0323)	0.0669** (0.0310)
Games Practice x Predicted Decline	0.0674** (0.0311)	0.0852*** (0.0326)	0.0628** (0.0315)
Math Practice	-0.0007 (0.0064)	-0.0048 (0.0068)	0.0016 (0.0059)
Games Practice	-0.0014 (0.0066)	-0.0045 (0.0070)	0.0010 (0.0061)
Dep. Var. Mean	0.47	0.47	0.47
Observations	329349	329349	329349

*Notes:* This table replicates Table II, Col. (2), using alternate approaches to calculate the “Predicted Decline” measure.

- “Predicted Decline” is defined as the difference in the percent of questions correct in the first quintile minus the given quintile. In Col. (1) we calculate this using baseline data from all students. We allow this to vary by school in Col. (2) (our preferred specification and the one shown in Table II, Col. (2)), and by school-test in Col (3).
- The other variable definitions, controls, and specifications are identical to that of Table II, Col. (2):
  - Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
  - “Cog. Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively.
  - All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.12: Treatment Effects on Performance Declines — Robustness to Long Ravens Tests

	Dependent Variable: 1[Question Correct]					
	Test Subject					
	All (1)	All (2)	Non-Math (3)	Math (4)	Listening (5)	Ravens (6)
<b>Panel A: Pooled Treatment Arms</b>						
Cog. Practice x Deciles 6-10	0.0114** (0.0047)					
Cog. Practice x Deciles 2-5	0.0074 (0.0047)					
Deciles 6-10	-0.0432*** (0.0035)					
Deciles 2-5	-0.0095*** (0.0033)					
Cog. Practice x Predicted decline		0.0854*** (0.0282)	0.0713** (0.0288)	0.1015** (0.0426)	0.0646** (0.0326)	0.0791* (0.0435)
Cog. Practice	-0.0011 (0.0060)	-0.0033 (0.0060)	-0.0002 (0.0062)	-0.0086 (0.0090)	-0.0012 (0.0066)	0.0002 (0.0096)
<b>Panel B: Disaggregated Treatment Arms</b>						
Math Practice x Deciles 6-10	0.0118** (0.0054)					
Games Practice x Deciles 6-10	0.0109** (0.0053)					
Math Practice x Deciles 2-5	0.0035 (0.0056)					
Games Practice x Deciles 2-5	0.0115** (0.0054)					
Math Practice x Predicted decline		0.0940*** (0.0322)	0.0942*** (0.0339)	0.0928* (0.0484)	0.0866** (0.0388)	0.1056** (0.0505)
Games Practice x Predicted decline		0.0769** (0.0330)	0.0479 (0.0335)	0.1113** (0.0500)	0.0424 (0.0380)	0.0517 (0.0507)
Math Practice	0.0001 (0.0067)	-0.0043 (0.0068)	-0.0054 (0.0071)	-0.0011 (0.0101)	-0.0051 (0.0077)	-0.0074 (0.0110)
Games Practice	-0.0023 (0.0069)	-0.0023 (0.0071)	0.0050 (0.0073)	-0.0164 (0.0106)	0.0029 (0.0078)	0.0081 (0.0112)
p-value: Math Decline = Games Decline		0.5984	0.1891	0.7077	0.2744	0.2980
Control Decline	0.12	0.12	0.05	0.18	0.06	0.03
Observations	368101	368101	167867	200234	66932	100935

*Notes:* This table replicates Table II, but includes questions from Ravens tests administered in one April follow-up round, which were excluded from the main analysis due to an error in test length that resulted in unreasonably long tests (e.g. up to 80 questions). In this table, we add the first  $x$  questions of the Ravens tests from the problematic round in the analysis sample;  $x$  equals the maximum number of questions in any Ravens exam across the other testing rounds (i.e. aside from the problematic round) for the student’s grade level.

- The variable definitions, controls, and specifications are identical to that of Table II:

- Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
- “Cog. Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively. “Deciles 2-5” and “Deciles 6-10” are binary indicators for if the question appears in the given decile range. “Predicted Decline” is defined at the item quintile-school level as the difference in the percent of questions correct in the first quintile minus the given quintile, and varies by school.
- Cols. (1) and (2)-(6) correspond to the specification in Equations 1 and 3, respectively. Cols. (1) and (2) estimate treatment effects for all three tests pooled. Cols. (3)-(6) show effects for the non-Math tests (listening and Ravens), then math, listening, and Ravens tests separately.
- All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.13: Treatment Effects on Performance Declines — Restricting to Attempted Questions Only

	Dependent Variable: 1[Question Correct]					
	Test Subject					
	All (1)	All (2)	Non-Math (3)	Math (4)	Listening (5)	Ravens (6)
<b>Panel A: Pooled Treatment Arms</b>						
Cog. Practice x Deciles 6-10	0.0134*** (0.0045)					
Cog. Practice x Deciles 2-5	0.0078 (0.0050)					
Cog. Practice x Predicted Decline		0.0894*** (0.0273)	0.0788*** (0.0282)	0.1036** (0.0425)	0.0643** (0.0323)	0.0977** (0.0451)
Cog. Practice	-0.0004 (0.0059)	-0.0021 (0.0059)	-0.0018 (0.0061)	-0.0020 (0.0090)	-0.0009 (0.0065)	-0.0039 (0.0100)
<b>Panel B: Disaggregated Treatment Arms</b>						
Math Practice x Deciles 6-10	0.0132** (0.0053)					
Games Practice x Deciles 6-10	0.0136*** (0.0051)					
Math Practice x Predicted Decline		0.0984*** (0.0315)	0.0979*** (0.0336)	0.0981** (0.0476)	0.0858** (0.0385)	0.1174** (0.0537)
Games Practice x Predicted Decline		0.0804** (0.0316)	0.0593* (0.0326)	0.1099** (0.0498)	0.0425 (0.0376)	0.0775 (0.0517)
Math Practice	0.0011 (0.0067)	-0.0030 (0.0068)	-0.0054 (0.0071)	0.0030 (0.0101)	-0.0046 (0.0075)	-0.0084 (0.0118)
Games Practice	-0.0019 (0.0069)	-0.0012 (0.0069)	0.0020 (0.0071)	-0.0072 (0.0106)	0.0028 (0.0077)	0.0007 (0.0115)
Dep. Var. Mean	0.55	0.55	0.51	0.58	0.55	0.47
Observations	279570	279570	128793	150777	66929	61864

*Notes:* This table replicates Table II, restricting to only question items in which the student provided an answer (i.e. the question was not left blank). The variable definitions, controls, and specifications are identical to that of Table II:

- Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
- “Cog. Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively. “Deciles 2-5” and “Deciles 6-10” are binary indicators for if the question appears in the given decile range. “Predicted Decline” is defined at the item quintile-school level as the difference in the percent of questions correct in the first quintile minus the given quintile, and varies by school.
- Cols. (1) and (2)-(6) correspond to the specification in Equations 1 and 3, respectively. Cols. (1) and (2) estimate treatment effects for all three tests pooled. Cols. (3)-(6) show effects for the non-Math tests (listening and Ravens), then math, listening, and Ravens tests separately.
- All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

Table A.14: Treatment Effects on Performance Declines — Robustness to Alternate Controls

	Dependent Variable: 1[Question Correct]					
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Pooled Treatment Arms</b>						
Cog. Practice x Deciles 6-10	0.0129*** (0.0047)	0.0127*** (0.0047)	0.0124*** (0.0047)			
Cog. Practice x Predicted Decline				0.0923*** (0.0286)	0.0899*** (0.0281)	0.0517** (0.0220)
<b>Panel B: Disaggregated Treatment Arms</b>						
Math Practice x Deciles 6-10	0.0127** (0.0055)	0.0125** (0.0055)	0.0121** (0.0055)			
Games Practice x Deciles 6-10	0.0131** (0.0054)	0.0130** (0.0054)	0.0127** (0.0054)			
Math Practice x Predicted Decline				0.0976*** (0.0330)	0.0947*** (0.0323)	0.0454* (0.0252)
Games Practice x Predicted Decline				0.0872*** (0.0330)	0.0852*** (0.0326)	0.0583** (0.0326)
Class Section FE	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes
Dep. Var. Mean	0.47	0.47	0.47	0.47	0.47	0.47
Observations	329349	329349	329349	329349	329349	329349

*Notes:* This table replicates Table II, Cols. (1) and (2), varying the set of controls included.

- Cols. (1) - (3) correspond to the specification in Table II, Col. (1). Cols. (4) - (6) correspond to the specification in Table II, Col. (2).
- Cols. (1) and (4) do not control for class section or student fixed effects. Col. (2) and (5) add in class section fixed effects (our preferred specification and the one included in Table II), and Col. (3) and (6) add student fixed effects.
- The other variable definitions, controls, and specifications are identical to that of Table II, Cols. (1) and (2):
  - Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
  - “Cog. Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively. “Deciles 6-10” is a binary indicator for if the question appears in the given decile range. “Predicted Decline” is defined at the item quintile-school level as the difference in the percent of questions correct in the first quintile minus the given quintile, and varies by school.
  - All regressions contain question and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.15: Treatment Effects on Declines — Robustness to Bootstrapping Standard Errors

	Dependent Variable: 1[Question Correct]				
	Test Subject				
	All (1)	Non-Math (2)	Math (3)	Listening (4)	Ravens (5)
<b>Panel A: Pooled Treatment Arms</b>					
Cog. Practice x Predicted Decline	0.0998 (0.0281)*** [0.0263]***	0.0790 (0.0281)*** [0.0281]***	0.1018 (0.0426)** [0.0333]***	0.0648 (0.0325)** [0.0316]**	0.0968 (0.0448)** [0.0367]**
Cog. Practice	-0.0046 (0.0060) [0.0060]	-0.0022 (0.0061) [0.0065]	-0.0084 (0.0090) [0.0081]	-0.0010 (0.0066) [0.0063]	-0.0046 (0.0100) [0.0098]
<b>Panel B: Disaggregated Treatment Arms</b>					
Math Practice x Predicted Decline	0.0947 (0.0323)*** [0.0313]***	0.0956 (0.0336)*** [0.0337]***	0.0935 (0.0484)* [0.0402]**	0.0869 (0.0387)** [0.0390]**	0.1109 (0.0535)** [0.0464]**
Games Practice x Predicted Decline	0.0852 (0.0326)*** [0.0297]***	0.0621 (0.0326)* [0.0326]*	0.1114 (0.0501)** [0.0406]***	0.0424 (0.0379) [0.0361]	0.0820 (0.0514) [0.0436]*
Math Practice	-0.0048 (0.0068) [0.0071]	-0.0060 (0.0071) [0.0078]	-0.0009 (0.0101) [0.0096]	-0.0050 (0.0076) [0.0076]	-0.0096 (0.0118) [0.0116]
Games Practice	-0.0045 (0.0070) [0.0068]	0.0019 (0.0071) [0.0075]	-0.0162 (0.0107) [0.0095]	0.0030 (0.0078) [0.0073]	0.0006 (0.0115) [0.0116]
Dep. Var. Mean	0.47	0.51	0.44	0.55	0.47
p-value: Math Decline = Games Decline	0.7601	0.3159	0.6989	0.2535	0.5910
Control Decline	0.12	0.05	0.18	0.06	0.03
Observations	329349	129115	200234	66932	62183

*Notes:* This table replicates Table II, Cols. (2)-(6), using bootstrapped standard errors.

- Standard errors are clustered by student and estimated using OLS (in parentheses), and they are bootstrapped (in brackets) to account for the fact that “Predicted Decline” is a constructed variable.
- The other variable definitions, controls, and specifications are identical to that of Table II:
  - Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
  - “Cog. Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively.
  - All columns correspond to the specification in Equation 3. Col. (1) estimates treatment effects for all three tests pooled. Cols. (2)-(5) show effects for the non-math tests (listening and Ravens), then math, listening, and Ravens tests separately.
  - All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.16: Treatment Effects on Mean Performance Across Duration of Declines Tests

	Dependent Variable: Avg. Score		
	Test Subject		
	Math (1)	Listening (2)	Ravens (3)
Cog. Practice	0.0073 (0.0081)	0.0097 (0.0060)	0.0106 (0.0085)
Observations	5541	5502	3649

*Notes:* This table tests for impacts on a student's average probability of getting a question correct, pooled across the duration of the declines tests.

- The regression specification corresponds to that Col. (7) of Table table: declines1, but estimates the average effect for each test subject separately.
- Cols. (1), (2), and (3) report average effects for the math, listening, and Ravens tests, respectively.
- All regressions control for class section and test version id fixed effects, and a linear control for the student's baseline average score. Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.17: Treatment Effects on Mean Performance on Math Test: Heterogeneous Effects by Question Difficulty

	Dependent Variable: Average Score			
	Heterogeneity Variable			
	ASER Questions	Easy Questions	Hard Questions	
	(1)	(2)	(3)	(4)
Math Practice	0.0106*	0.0102*	0.0113*	0.0103*
	(0.0057)	(0.0056)	(0.0062)	(0.0062)
Games Practice	0.0073	0.0071	0.0071	0.0065
	(0.0059)	(0.0058)	(0.0063)	(0.0063)
Hte. Variable		0.0060	0.0061***	-0.0027**
		(0.0048)	(0.0014)	(0.0013)
Math Practice x Hte. Variable		-0.0037	-0.0029	0.0015
		(0.0083)	(0.0022)	(0.0021)
Games Practice x Hte. Variable		-0.0118	-0.0020	0.0005
		(0.0080)	(0.0022)	(0.0021)
F-test p-value Total Math Het. Effect		0.5286	0.1808	0.0568
F-test p-value Total Games Het. Effect		0.6364	0.4261	0.2625
Observations	14692	17851	48211	48211

*Notes:* This table examines students' average probability of getting a question correct on the math declines test, testing for heterogeneity in effects by question difficulty.

- Observations are at the student-test-difficulty bin level. The dependent variable is the proportion of questions a student answered correctly, averaged across questions in the given student-test-difficulty bin.
- Col. (1) shows overall effects by treatment arm. The "Heterogeneity variable" varies by column, and corresponds to the question difficulty bin. In Col. (2), it is a binary variable that indicates questions similar to those administered on Annual Status of Education Report (ASER) exams, which are basic skills questions. In Col. (3) it indicates "easy" questions, defined as those in the bottom quartile of question difficulty as measured by question fixed effects. In Col. (4) it indicates "hard" questions, defined as those in the top quartile of question difficulty as measured by question fixed effects.
- "Math Practice" and "Games Practice" denote the Math or Games sub-treatments, respectively.
- \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Table A.18: Treatment Effects on Performance Declines by Testing Round

	Dependent Variable: 1[Question correct]			
	Definition of Treat Variable			
	Cognitive Practice	Math Practice	Games Practice	Cognitive Practice
	(1)	(2)	(3)	(4)
Cog. Practice x Predicted decline x Dec.	0.0863** (0.0336)	0.0857** (0.0388)	0.0536 (0.0379)	
Cog. Practice x Predicted decline x Feb.	0.0991*** (0.0361)	0.1240*** (0.0478)	0.1018** (0.0482)	
Cog. Practice x Predicted decline x Follow-up	0.0861** (0.0426)	0.0859* (0.0499)	0.0859* (0.0489)	
Cog. Practice x Deciles 6-10 x Dec.				0.0394*** (0.0058)
Cog. Practice x Deciles 6-10 x Feb.				-0.0083 (0.0064)
Cog. Practice x Deciles 6-10 x Follow-up				0.0097 (0.0101)
F-test p-value Dec. = Feb.	0.7117	0.4749	0.3584	0.0000
F-test p-value Feb. = Follow-up	0.7766	0.4966	0.7779	0.1298
Dep. Var. Mean	0.47	0.47	0.47	0.46
Observations	329349	219341	217223	329349

*Notes:* This table examines the treatment effect on performance declines in the listening, Ravens and math exams by testing round. The table is similar to the structure of Table table: persistence, but with the "Endline" tests further disaggregated by the Dec (midline) and Feb (endline) testing rounds.

- Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
- "Cog. Practice" denotes receiving a treatment, where the column header describes which treatment arm is being tested. "Dec", "Feb" and "Follow-up" are binary indicators that equals one if the test is a midline (conducted in December), endline (conducted in February) or follow-up test (administered roughly 3 to 5 months after the end of the intervention). "Predicted Decline" is defined at the item quintile-school level as the difference in the percent of questions correct in the first quintile minus the given quintile, and varies by school. "Deciles 6-10" is a binary indicator that equals one if the question appears in the second half of the test.
- Cols. (1) and (4) pool both sub-treatments, and Cols. (2) and (3) present the "Math Practice" and "Games Practice" separately.
- For ease of interpretation, coefficients presented as total effects rather than marginal effects. An F-test of equality of coefficients is presented at the bottom of the table.
- All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.19: SART and Symbol Matching Results by True and False Positive Rates

	Dependent Variable: Z-score			
	Test Subject			
	Pooled (1)	SART (2)	Symbol Matching (3)	Pooled (4)
<b>Panel A: True Positive Rate</b>				
Cognitive Practice	0.0476* (0.0254)	0.0634* (0.0343)	0.0364 (0.0321)	
<i>Sub-treatments:</i>				
Math Practice				0.0701** (0.0288)
Games Practice				0.0248 (0.0299)
p-value: Math Practice = Games Practice				0.1246
<b>Panel B: False Positive Rate</b>				
Cognitive Practice	-0.0303 (0.0236)	-0.0562* (0.0330)	-0.0131 (0.0317)	
<i>Sub-treatments:</i>				
Math Practice				-0.0106 (0.0292)
Games Practice				-0.0503* (0.0263)
p-value: Math Practice = Games Practice				0.1773
Observations	9699	3895	5804	9699

*Notes:* This table provides additional detail on the treatment effects on two measures of sustained attention drawn from the psychology literature.

- Observations are at the student-test level. Panel A examines the student's true-positive z-score, winsorized at the 99th percentile. Panel B examines the student's false-positive z-score, winsorized at the 99th percentile.
- Experimental tasks are denoted at the top of each column. The Sustained Attention to Response Task (SART) is a computer-based task, which measures sustained focus via reaction times to stimuli. Symbol Matching is a paper and pencil-based task in which students search through a grid of symbols to identify the target symbol and repeat the search for a series of unique grids over 15 minutes.
- Cols. (1) and (4) pool across the two tasks. Cols. (2) and (3) present the SART and symbol matching results separately.
- "Cognitive Practice" denotes receiving either treatment. "Math Practice" and "Games Practice" denote the Math or Games sub-treatments, respectively.
- All regressions contain class section fixed effects and a linear control for the student's average baseline performance on the two tasks. Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.20: Treatment Effects on Performance Declines — Restricting to Observations without Performance Incentives

	Dependent Variable: 1[Question Correct]					
	Test Subject					
	All (1)	All (2)	Non-Math (3)	Math (4)	Listening (5)	Ravens (6)
<b>Panel A: Pooled Treatment Arms</b>						
Cog. Practice x Deciles 6-10	0.0124** (0.0048)					
Cog. Practice x Deciles 2-5	0.0079* (0.0048)					
Deciles 6-10	-0.0437*** (0.0037)					
Deciles 2-5	-0.0112*** (0.0036)					
Cog. Practice x Predicted decline		0.0921*** (0.0293)	0.0801*** (0.0292)	0.1043** (0.0436)	0.0721** (0.0345)	0.0918** (0.0455)
Cog. Practice	-0.0011 (0.0060)	-0.0037 (0.0062)	-0.0007 (0.0063)	-0.0085 (0.0093)	0.0014 (0.0068)	-0.0037 (0.0100)
<b>Panel B: Disaggregated Treatment Arms</b>						
Math Practice x Deciles 6-10	0.0122** (0.0055)					
Games Practice x Deciles 6-10	0.0126** (0.0055)					
Math Practice x Deciles 2-5	0.0032 (0.0056)					
Games Practice x Deciles 2-5	0.0127** (0.0055)					
Math Practice x Predicted decline		0.0950*** (0.0337)	0.0957*** (0.0347)	0.0938* (0.0496)	0.0963** (0.0406)	0.1007* (0.0540)
Games Practice x Predicted decline		0.0897*** (0.0338)	0.0641* (0.0338)	0.1161** (0.0508)	0.0472 (0.0401)	0.0821 (0.0525)
Math Practice	0.0010 (0.0068)	-0.0036 (0.0070)	-0.0051 (0.0073)	0.0009 (0.0105)	-0.0039 (0.0077)	-0.0083 (0.0118)
Games Practice	-0.0033 (0.0070)	-0.0038 (0.0072)	0.0038 (0.0073)	-0.0182* (0.0109)	0.0069 (0.0079)	0.0010 (0.0117)
p-value: Math Decline = Games Decline		0.8757	0.3791	0.6556	0.2443	0.7374
Control Decline	0.12	0.12	0.05	0.18	0.06	0.03
Observations	325892	325892	125658	200234	64083	61575

*Notes:* This table replicates Table II, restricting to students who did not receive incentives. The variable definitions, controls, and specifications are identical to that of Table II:

- Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.

- “Cog. Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively. “Deciles 2-5” and “Deciles 6-10” are binary indicators for if the question appears in the given decile range. “Predicted Decline” is defined at the item quintile-school level as the difference in the percent of questions correct in the first quintile minus the given quintile, and varies by school.

- Cols. (1) and (2)-(6) correspond to the specification in Equations 1 and 3, respectively. Cols. (1) and (2) estimate treatment effects for all three tests pooled. Cols. (3)-(6) show effects for the non-Math tests (listening and Ravens), then math, listening, and Ravens tests separately.

- All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.21: Impact of Treatment on Discouragement Effects

<i>Question Difficulty:</i>	Dependent Variable: 1[Question Correct]				
	All Tests				Listening Tests
	(1)	Top Decile (2)	Top Quartile (3)	Top Half (4)	Top Quartile (5)
Cognitive Practice	-0.0024 (0.0063)	-0.0028 (0.0063)	-0.0030 (0.0067)	-0.0050 (0.0068)	0.0022 (0.0075)
Cognitive Practice x Predicted Decline	0.0832*** (0.0318)	0.0829*** (0.0318)	0.0828*** (0.0318)	0.0819** (0.0318)	0.0525 (0.0354)
Lagged Hard Question		-0.0129** (0.0063)	-0.0200*** (0.0047)	-0.0069* (0.0318)	-0.0257*** (0.0354)
Cognitive Practice x Lagged Hard Question		0.0067 (0.0065)	0.0025 (0.0050)	0.0055 (0.0039)	0.0015 (0.0082)
Observations	314892	314892	314892	314892	61607

*Notes:* This table shows the effect of treatment on whether the student gets the question *after* a difficult question correct. Question difficulty is defined at the top of each column.

- Observations are at the student-test-question level. Question item order was randomized across students. The dependent variable is a binary indicator for whether the question is correct.
- “Lagged Hard Question” is an indicator that equals one if the question comes immediately after a difficult question. Column 1 replicates Panel A Column 2 of Table II, but has fewer observations because “Lagged Hard Question” is not defined for the first question. “Cognitive Practice” denotes receiving either treatment. “Predicted Decline” is defined at the item quintile-school level as the difference in the percent of questions correct in the first quintile minus the given quintile, and varies by school.
- “Top Decile”, “Top Quartile”, and “Top Half” indicate the definition of question difficulty used in each column, and refer to the bottom 10%, 25%, and 50% questions in terms of accuracy, respectively, computed using control group data.
- All regressions contain question, class section, and test version fixed effects, a linear control for baseline average score, and a linear control for the fraction of students in the same school who got the question correct (computed using control group data only). Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.22: Treatment Effects on Attendance

	<b>Dependent Variable: Absent on main testing day</b>	
	(1)	(2)
Cognitive Practice	-0.0045 (0.0047)	
<i>Sub-treatments:</i>		
Math Practice		-0.0066 (0.0053)
Games Practice		-0.0024 (0.0054)
Observations	13635	13635

*Notes:* This table examines whether there is an effect of the treatment on the likelihood a student is absent on a given day.

- Observations are at the student-day level. We examine attendance on the days of the (unannounced) tests administered to students as part of the experiment post-baseline.
- Columns (1) and (2) pool across all five tests (listening, math, Ravens, SART and COS) and all testing rounds.
- The dependent variable is a binary indicator for whether the child was absent on that day of school. A student is identified as absent if they are not present at school on the day the rest of their class takes the test and instead completes the test on a make-up test day at some point afterward.
- “Cognitive Practice” denotes receiving either treatment. “Math Practice” and “Games Practice” denote the Math or Games sub-treatments, respectively.
- Standard errors are clustered by student. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.23: Treatment Effects on School Performance — Heterogeneity by Subjectivity in Grading

	Dependent Variable: Z-score				
	Heterogeneity Group				
	English & Hindi	Grades 3-5	Hindi & English or Grades 3-5	Hindi & English and Grades 3-5	Non-chain Schools
	(1)	(2)	(3)	(4)	(5)
<b>Panel A: All Schools</b>					
Cog. Practice x Het. Group	0.0106 (0.0345)	0.0198 (0.0690)	-0.0211 (0.0555)	0.0419 (0.0594)	0.0000 (0.0000)
Cog. Practice	0.0826** (0.0381)	0.0787 (0.0520)	0.1078* (0.0577)	0.0742* (0.0394)	0.0897** (0.0348)
Het. Group	0.0491* (0.0285)	0.0330 (0.0575)	0.0374 (0.0471)	0.0609 (0.0488)	0.0000 (0.0000)
<b>Panel B: Chain Schools Only</b>					
Cog. Practice x Het. Group	0.0178 (0.0485)	0.0063 (0.0759)	-0.0282 (0.0646)	0.0402 (0.0748)	
Cog. Practice	0.1087** (0.0446)	0.1188** (0.0595)	0.1451** (0.0688)	0.1074** (0.0457)	
Het. Group	0.0784* (0.0400)	0.1045* (0.0619)	0.0725 (0.0540)	0.1423** (0.0604)	
Observations	11320	11320	11320	11320	11320

*Notes:* This table examines whether treatment effects on school grades vary according to how subjective or objective the grading of the course is.

- Panel A examines all schools in our sample. Panel B includes only the schools in a chain with highly standardized grading practices.
- This table mirrors Table I Panel A, but adds an interaction between treatment and the covariate denoted at the top of the column.
- “Covariate” varies by column and is intended to capture more subjective grading practices in a variety of ways. In Col. (1) is covariate captures whether the grade is in either Hindi or English (rather than math). In Col. (2) it captures whether the student is currently in grades 3-5 (inclusive), where test items become more complex (e.g. essays rather than fill in the blanks). Col. (3) is an indicator if the grade was in Hindi or English *or* the student is in grades 3-5. Col. (4) is an indicator which takes the value 1 if the grade was in Hindi or English *and* the student is in grades 3-5. Col. (5) is an indicator for non-chain schools.
- The other variable definitions, controls, and specifications are identical to that of Table I Panel A. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

Table A.24: Impact of Treatment on Behaviors Outside of School

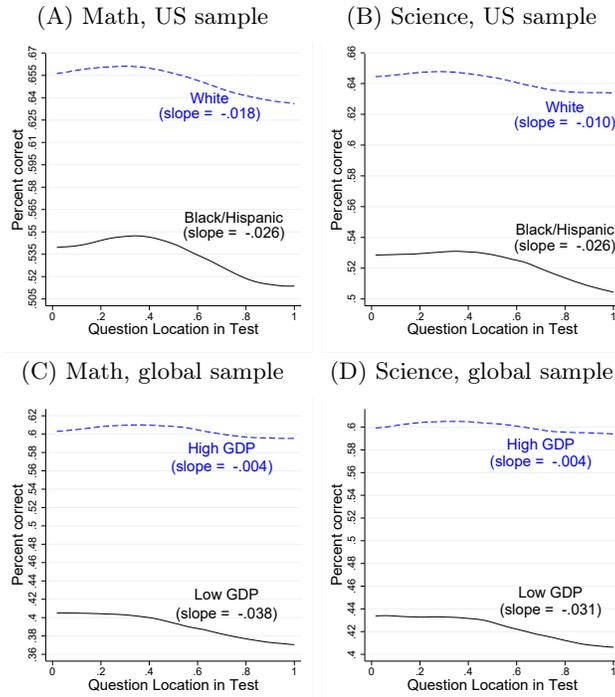
	(1) Breakfast	(2) Practice	(3) HW Time	(4) HW Help
Cognitive Practice	-0.0346 (0.0623)	0.0766 (0.0497)	-0.0280 (0.0586)	-0.0317 (0.0623)
Dep. Var. Mean	1.656	1.591	1.603	1.055
Dep. Var. SD	0.760	0.620	0.743	0.735
Observations	706	706	706	706

*Notes:* This table shows the effect of treatment assignment on students' experience outside of school.

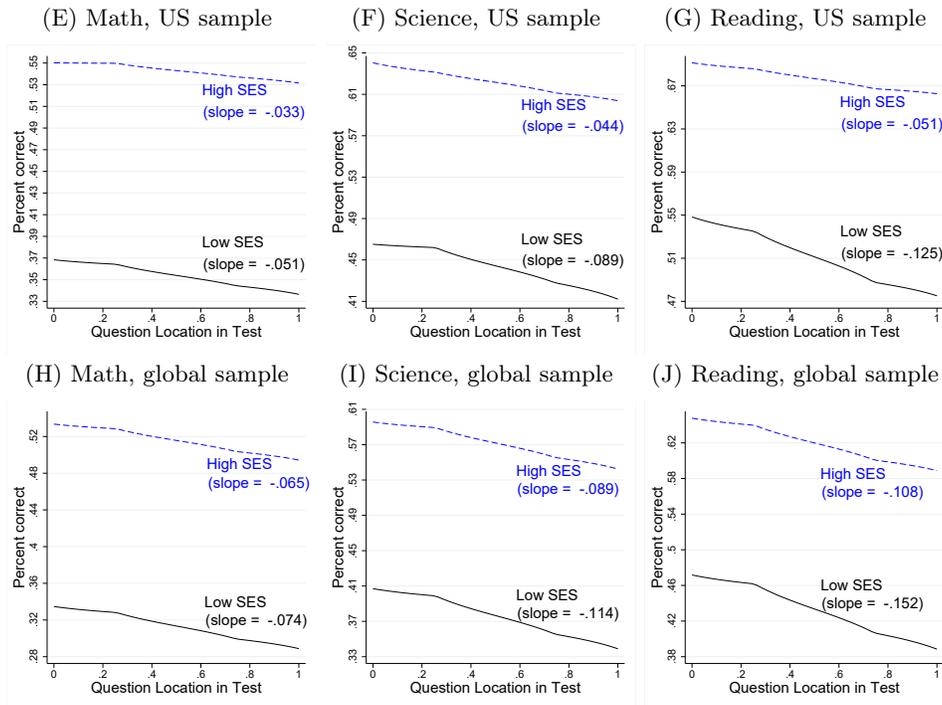
- Observations are at the student level. The dependent variable is student self-report of conditions at their home along different criteria (denoted at the top of the column).
- "Breakfast" is a measure of the total number of items a student had for breakfast that day from the following list: egg, bread, rice, paratha, cereal, milk, tea, fruit, meat, and other. "Practice" is a count of the number of cognitively-focused practice activities from a pre-specified list a child did after school, ranging from 0-3. "HW Time" is a measure for how much time students spend on homework scored as 0 (less than 45 minutes), 1 (about 45 minutes), or 2 (more than 45 minutes). "HW Help" captures the number of individuals (family members, tutor, etc) that help them with their homework.
- "Cognitive Practice" is dummy for whether the student was assigned to the treatment.
- All regressions contain class section fixed effects. Robust standard errors are reported in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**FIGURE A.1: Performance Declines in Achievement Tests**

**TIMSS Exam**



**PISA Exam**



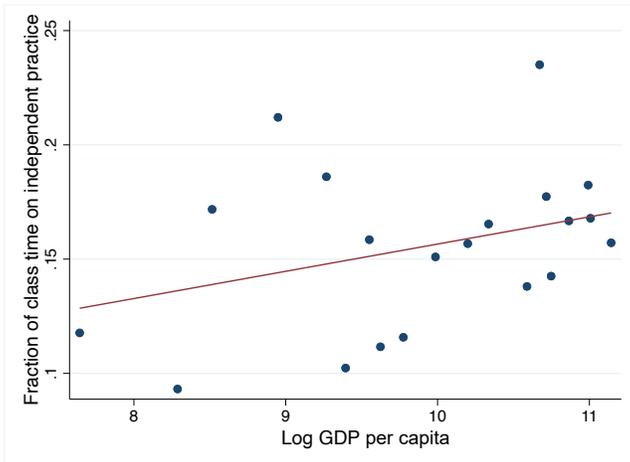
*Notes:* The figures show student performance over the length of the TIMSS and PISA tests on a single y-axis. All other variable definitions, controls, and specifications are identical to that of Figure I.

- TIMSS (PISA) is administered to fourth graders (15 year-olds) in more than 50 countries. All subjects administered are presented here. Observations are at the student-question level.
- For TIMSS, question order is block randomized within each test subject; graphs plot residuals after removing question fixed effects. For PISA, randomization is across test subjects (with 4 randomization blocks per exam); we remove question block fixed effects.
- The x-axis is “Question location in test”, which denotes where in the exam the question item appeared normalized on a scale of 0 to 1 (i.e. question number within subject in TIMSS and question block number across the exam in PISA). The y-axis plots the average score (i.e. percent answered correctly) for each question location on the test.
- The plots display the smoothed values of a kernel-weighted local polynomial regression, with a bandwidth of 0.15 for TIMSS and larger bandwidth of 0.33 for PISA (due to the smaller number of randomization blocks).
- In the TIMSS US sample (A-B), relative advantage is proxied by race (white and non-white, respectively). In the TIMSS global sample, these differences are proxied by the top (bottom) quartile of GDP/capita (C-D). In the PISA data (E-J), high (low) SES is proxied by the top (bottom) quartile of the ESCS measure, an index capturing parental income, occupation, and education.

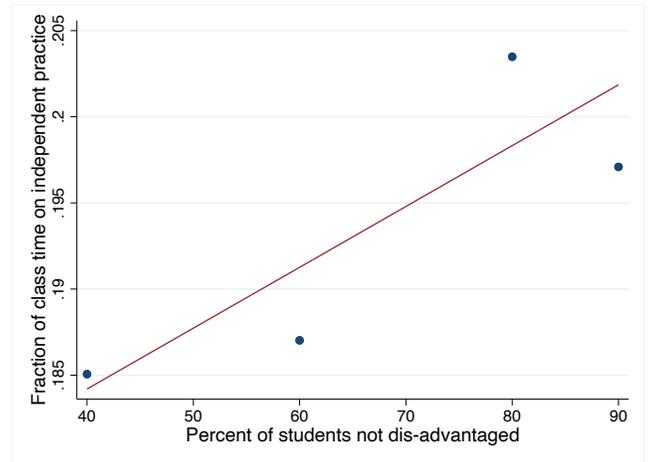
**FIGURE A.2:** Differences in Schooling Practices by Social and Economic Advantage

**Frequency of Independent Practice**

(A) Global sample

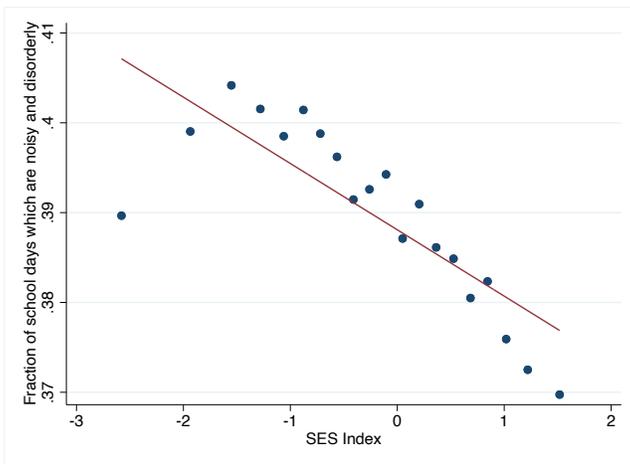


(B) US sample

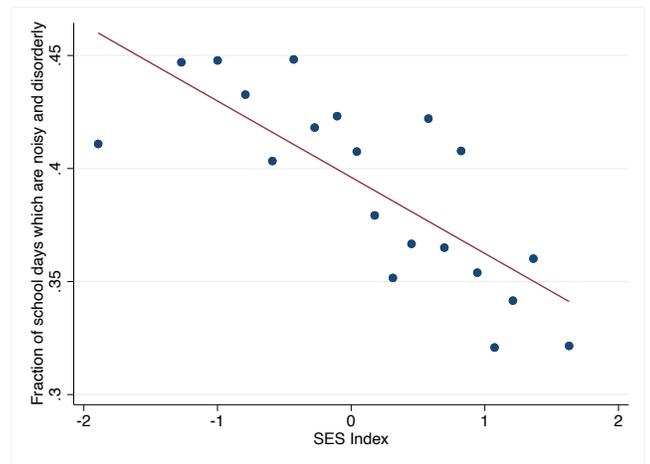


**Frequency of Classroom Distractions**

(C) Global sample



(D) US sample



*Notes:* The figures show the relationship between income and schooling environment.

- Panels A and B present data from the TIMSS teacher survey on pedagogy used within the classroom. The y-axis is the fraction of class time spent on independent practice. Teachers rate how often students engage in this type of activity on a 4-pt scale from “never” (coded as 0) to “every or almost every lesson” (coded as 0.75). In Panel A, the sample is all countries, and the x-axis is log GDP per capita. In Panel B, the sample is the US, and the x-axis is the percent of students within the school who are not disadvantaged (where the fraction of disadvantaged students is reported by school administrators from among 4 discrete options).
- Panels C and D present data from the PISA teacher survey. The x-axis is a student-level SES index constructed by PISA. The y-axis is the fraction of classes in which there is noise and disorder, rated by teachers on a 4-pt scale from “never” (coded as 0) to “every lesson” (coded as 1). The data is grouped into ventiles, presenting the average within each ventile (blue dots).
- In each plot, the red line is the line of best-fit.

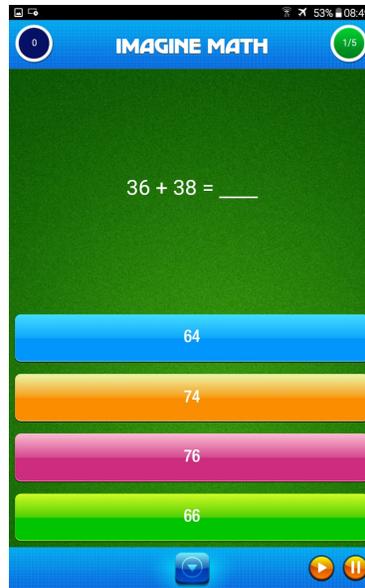
**FIGURE A.3:** Example Classrooms from Study Schools



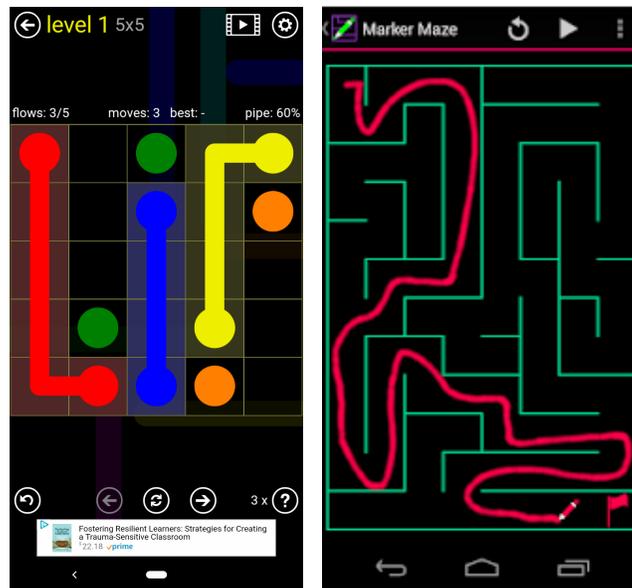
*Notes:* The photographs each show a classrooms from two of our study schools to provide context.

**FIGURE A.4:** Treatment Tablet Software

(a) Math practice treatment



(b) Games practice treatment



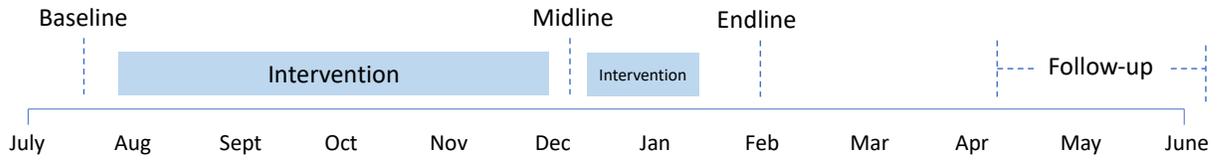
*Notes:* These figures show example screenshots from the treatment tablet software used throughout the intervention. For the Math Practice, we use the imagineMath software, developed by Pixatel. For the Games Practice, we use simple games with limited animation downloaded from the Android app store.

**FIGURE A.5:** Program Treatment Classes



*Notes:* The photographs show two example treatment program classes.

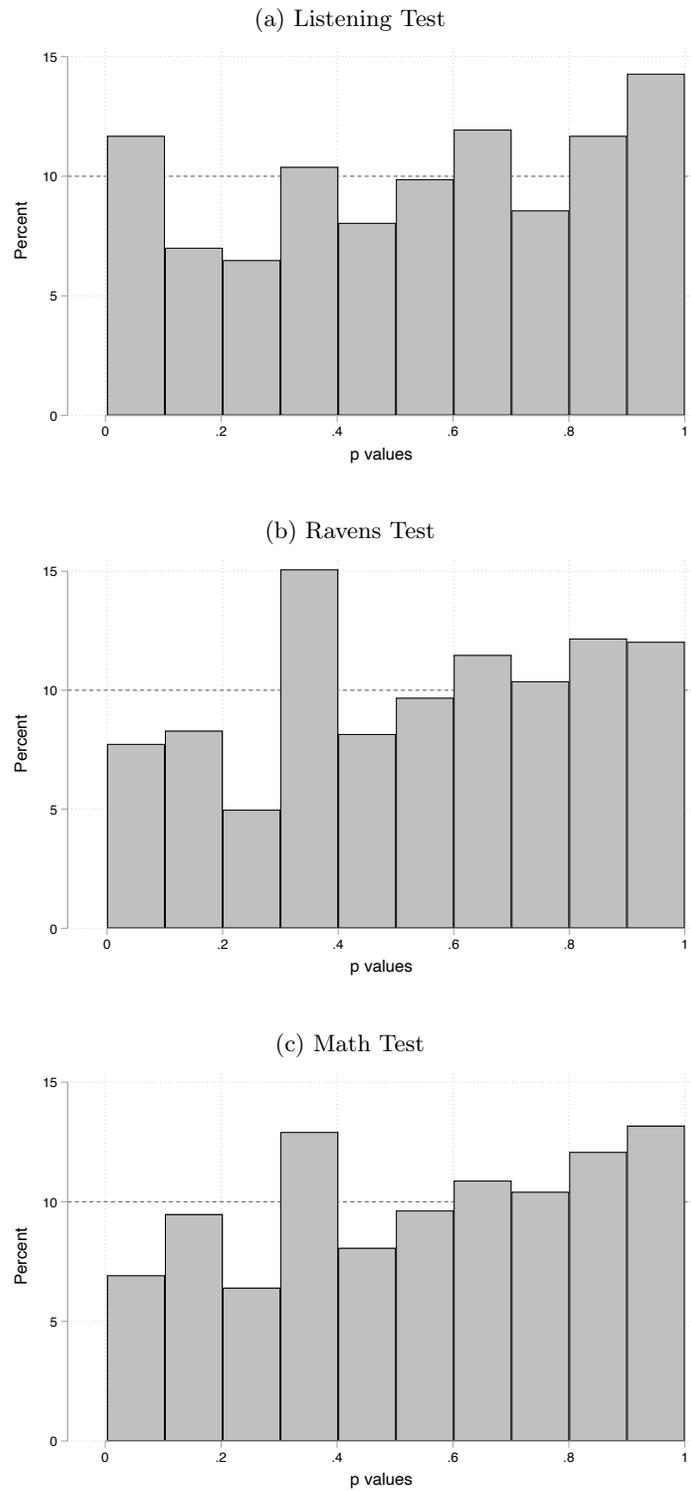
**FIGURE A.6:** Experiment timeline



*Notes:* This figure shows the timeline of the intervention for a typical school. Program treatment and control classes were administered from August to early December and again in January. Baseline tests were conducted in July and August (before the start of program classes). Midline tests were conducted during the intervention break in December, and endline tests were conducted in early February. Follow-up tests were conducted from late April through June.

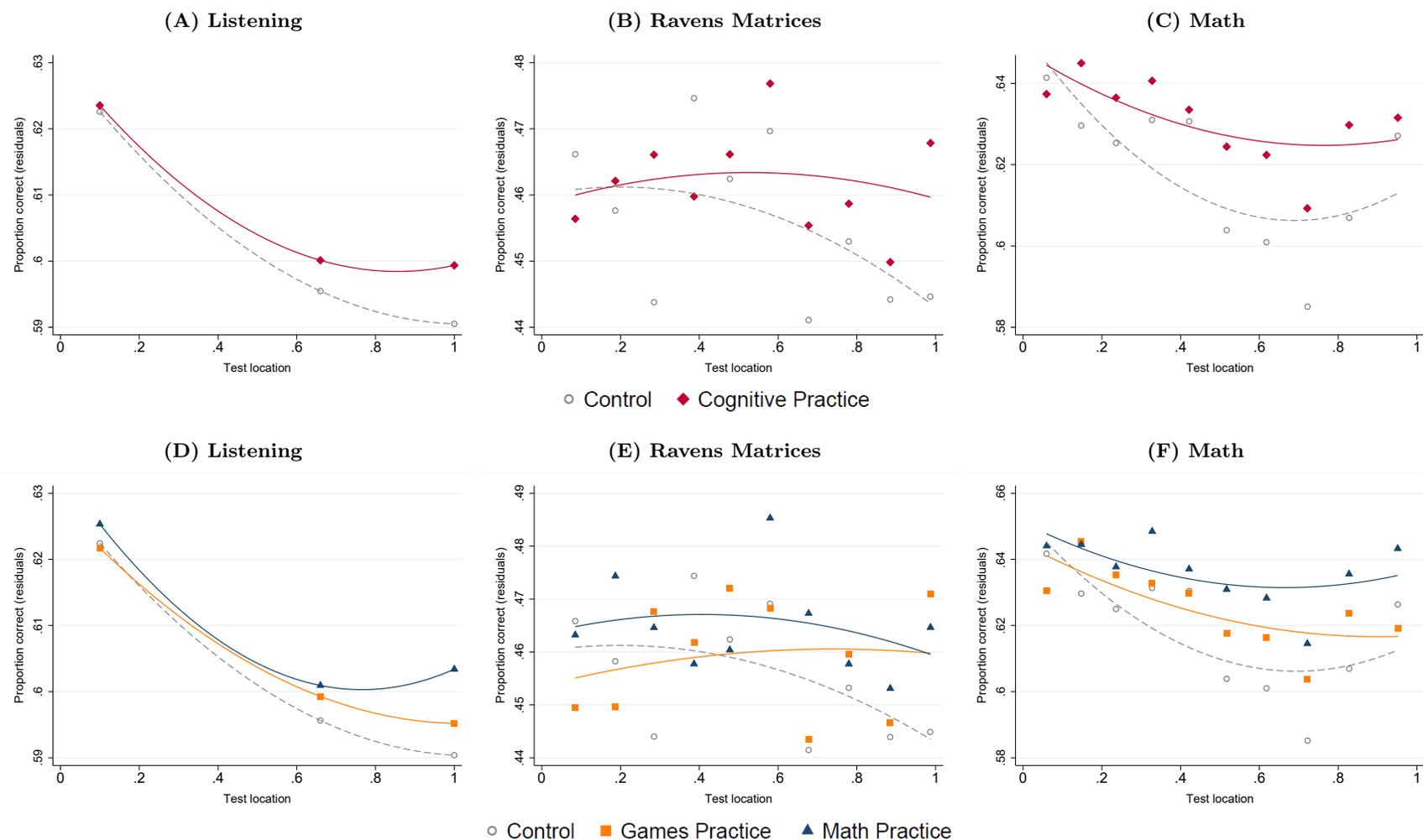


**FIGURE A.8:** Randomization Balance of Test Versions



*Notes:* These plots show the distribution of p-values for 4,478 coefficients of whether the student received a given test version on dummies for treatment status. These regressions are calculated within each test-round-school-grade. This tests whether the allocation of test packet versions is balanced across treatment status. For a perfectly random allocation of test versions, as the sample size goes to infinity, we would expect each bar to approach 10%.

**FIGURE A.9:** Attempted Questions Only: Performance Over the Length of the Test by Treatment

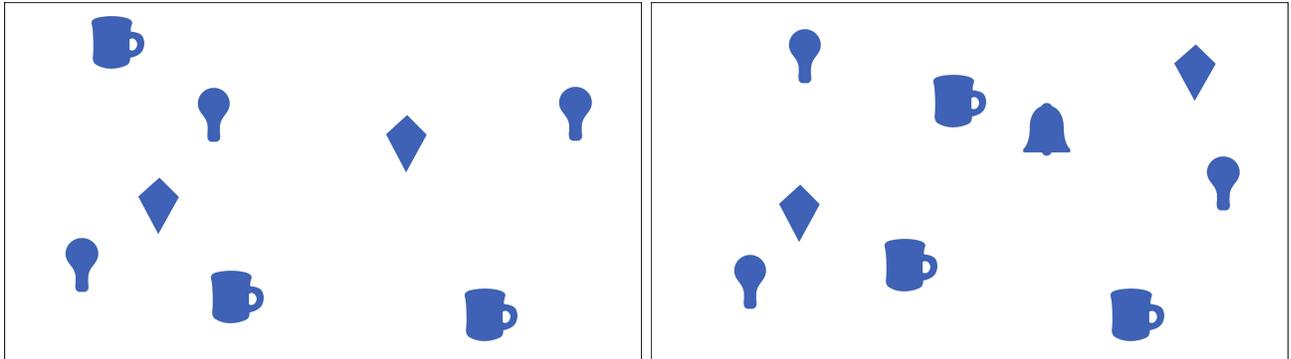


*Notes:* The figures replicate the decline plots on the listening, Raven’s Matrices and math tests from Figure II, but restrict to questions students attempted (i.e. the student did not leave the question blank). All variable definitions and controls are identical to that of Figure II:

- Question order is randomized in each test. Observations are at the student-test-question level.
- Each panel displays a binscatter plot, where the x-axis is the percent location of the test, grouped in deciles (where 0.1 is the first decile of the test and 1 is the last decile), and the y-axis is the proportion of students who answer the question correctly. Data is residualized to remove question and test version fixed effects. All plots are overlaid with a quadratic best-fit curve by experimental arm.
- The top row compares the control group (dashed gray, open circles) to the pooled sub-treatments (solid red, diamonds). The bottom row compares the Control (dashed gray, open circles) to each sub-treatment – Math arm (solid blue, triangles) and Games arm (solid orange, squares) – individually.
- For ease of interpretation of decline magnitudes, in each plot, the decile 1 control group mean is added to all residuals.
- Table A.13 presents the full set of corresponding treatment effects estimates.

**FIGURE A.10:** Example Test Packet and Photos - SART and Symbol Matching Task

(a) Screenshots from SART software



(b) Symbol Matching Example Test Page

i +														
+	Γ	+	∅	“	+	ι	⊕	Γ	+	+	≡	κ	∅	“
+	Γ	∩	∞	∞	ι	+	⊙	+	Γ	∩	ι	∩	ι	κ
∕	≡	Γ	∩	κ	∞	ι	∅	Γ	ι	∞	∕	+	+	∩
∫	≡	∞	+	∞	∞	+	≡	∞	+	≡	∞	∞	∞	+
∩	Γ	≡	∞	ι	∕	+	∞	κ	∕	Γ	∩	Γ	+	+
ι	∩	+	∞	∞	∅	Γ	⊙	≡	+	∞	∅	+	∞	+
∩	Γ	Γ	κ	⊙	≡	∞	∅	∞	ι	κ	∞	∩	+	∩
∩	∞	∞	ι	∩	∞	∞	∞	+	κ	∞	≡	∞	∩	∞
≡	∞	∩	∞	+	Γ	κ	∞	ι	∕	≡	+	κ	∞	≡
+	κ	∞	ι	∅	∅	ι	∞	≡	∞	∞	∞	∞	∞	≡

*Notes:* These figures show information about the two measures of attention from psychology, SART and the symbol matching task.

- Panel A shows two screenshots of the SART software. The left hand panel shows the software when the target stimulus – a bell – is not present. The right hand panel has the target stimulus present, indicating the students should press the space bar as quickly as possible.
- Panel B shows an example test page from the symbol matching task with the target symbols shown at the top of the page. Students needed to find the matching symbols in the grid below.

FIGURE A.11: Example Test Pages - Grade 4 Math and English exams

(a) First page of Math Exam

**HALF-YEARLY EXAMINATION 2018-19**  
**MATHEMATICS**  
**CLASS IV**

Two hours Maximum marks : 70

---

**Instructions:**  
Please do not write anything on your question paper except your name and roll number.  
Write your name very neatly and clearly on your answer script.  
Answers to this paper must be written in neat and legible handwriting on the answer script provided separately.  
Do not copy the questions on your answer script. Write the correct question number only.

---

**Section A [30 marks]**  
Attempt all questions from this section.

**Question 1**  
Fill in the blanks: (Write the answers only)

(a) The predecessor of 2000 is \_\_\_\_\_. [1]

(b)  $63,179 - \underline{\hspace{2cm}} = 63,179$  [1]

(c)  $555 \times 1000 = \underline{\hspace{2cm}}$  [1]

(d)  $14,128 \div 14,128 = \underline{\hspace{2cm}}$  [1]

(e) The fractions with same denominator are called \_\_\_\_\_ fractions. [1]

**Question 2**

(a) Write the face value and place value of 8 in the number 2,81,657. [2]

---

- 6      This paper consists of five printed pages.      Turn over.

(b) First page of English Exam

**HALF-YEARLY EXAMINATION 2018-19**  
**ENGLISH LITERATURE**  
**CLASS IV**

Two hours Maximum marks : 80

Spoken English + Elocution : 5+5  
Reading : 10  
Dictation : 5  
Handwriting : 5  
Textbook : 50

---

**Instructions:**  
Please do not write anything on your question paper except your name and roll number.  
Write your name very neatly and clearly on your answer script.  
Answers to this paper must be written in neat and legible handwriting on the answer script provided separately. **You will be marked out of 5 marks for your handwriting.**  
Do not copy the questions on your answer script. Write the correct question number only.

---

**Question 1**  
Answer any four of the following questions:

(a) What did the elephant look like in far-off times? [3]

(b) What was so fascinating about Gobstoppers? [3]

(c) What story of her life did Avva narrate to her granddaughter? (Any three points) [3]

(d) Where was the Elephant's Child going to find out what the crocodile has for dinner? What did he take with him? [3]

(e) Describe the filth that clung around Mrs Pratchett. [3]

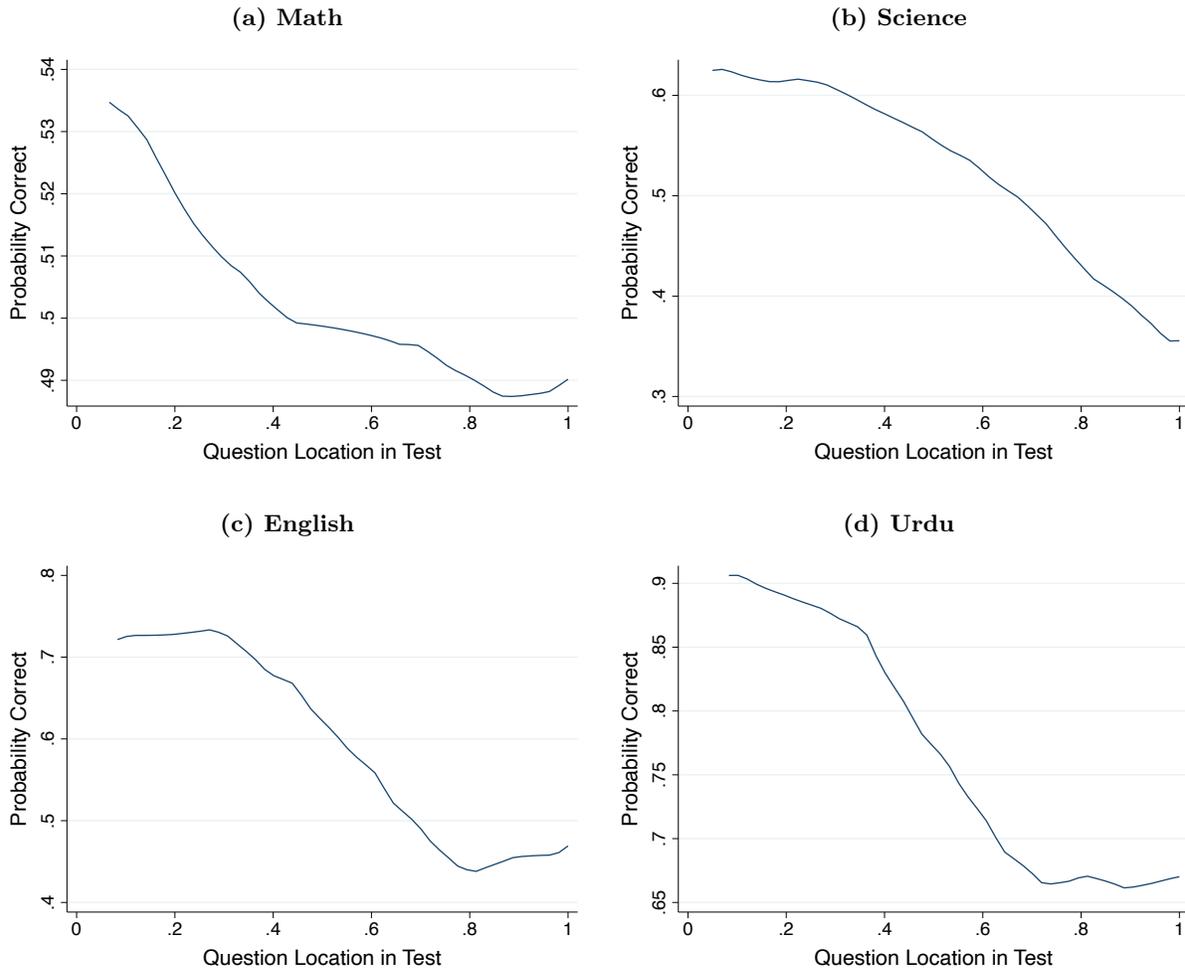
(f) What happened when the Dassara festival came? [3]

---

- 3      This paper consists of four printed pages.      Turn over

Notes: These figures show the first page of school administered exams in Math (Panel (a)) and English (Panel (b)) for 4th grade students.

**FIGURE A.12:** Supplementary Data: Performance Over the Length of the Test by Subject

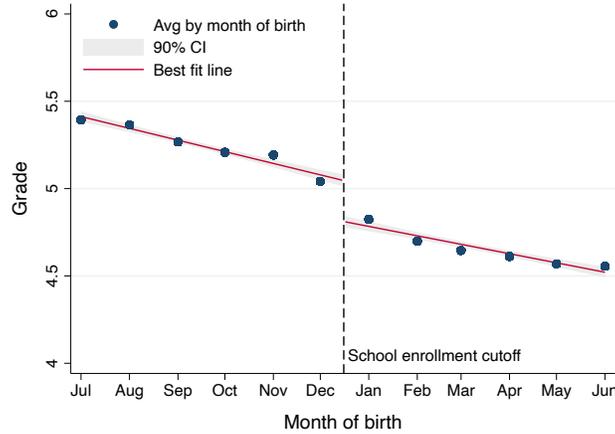


*Notes:* The figures show student performance over the length of the test by test subject using a sample of 5,353 9-11 year-olds in Pakistan from Brown and Andrabi (2021).

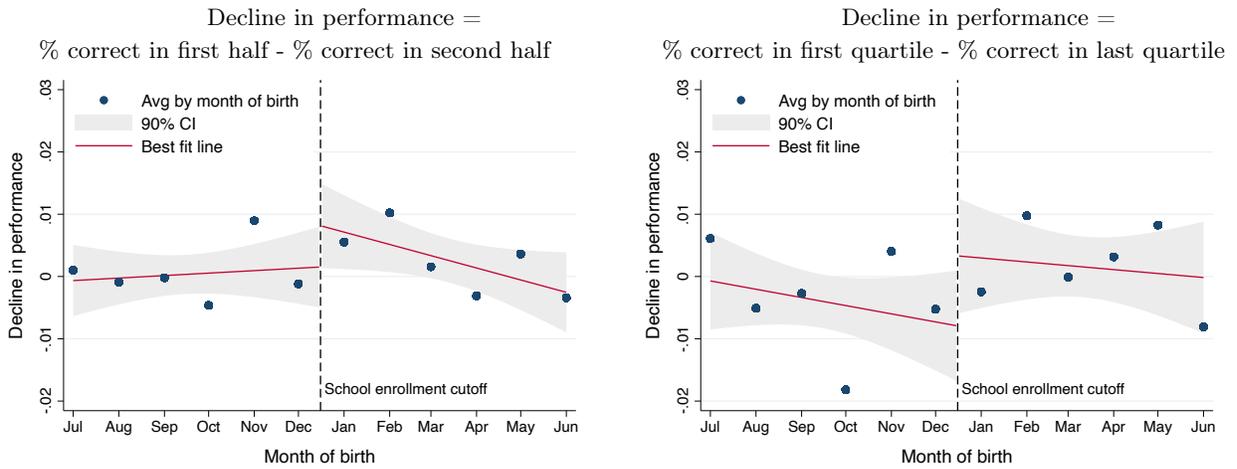
- Observations are at the student-test-question level. Question order on the tests is randomized.
- The figure plots the probability a question was answered correctly (y-axis) against the percent location of the question on the test (where 0 is the beginning of the test and 1 is the end of the test, x-axis).  $N = 217,516$ .

**FIGURE A.13:** Effect of Additional Schooling on Performance Decline

**(a) First stage: Students' grade-level by birth month**



**(b) Reduced form: Decline in performance by birth month**



*Notes:* The figures show students' average grade-level and performance decline by month of birth using a sample of 5,353 9-11 year-olds in Pakistan from Brown and Andrabi (2021).

- The kindergarten enrollment cut-off is December 31, shown by the vertical dashed line. The figure allows for differential slope on either side of the cutoff. Question order on the tests is randomized.
- *Panel A:* The figure plots students' average grade-level at the time of the exam by month of birth. Children born in January have 0.22 fewer years of schooling, conditional on their age.
- *Panel B:* The two figures plot students' decline over the test (performance in the beginning minus performance at the end) by the child's month of birth. Students' who are just to the right of the cutoff (and therefore have fewer years of schooling) have significantly more performance decline, conditional on age.
- Performance decline is captured using two metrics. In the left figure of Panel B, performance decline is defined as the average percent correct on items in the first half of the test minus the average percent correct on items in the second half of the test. The right figure is similar but uses the difference between the average percent correct in the first quartile minus the last quartile of the test.

## B Data and Analysis Appendix

### Additional Details of Regression Discontinuity in Section

In section 7 we present the results of the effect of an additional school year on cognitive endurance. We instrument for years of schooling by the predicted years of schooling using a regression discontinuity approach at the enrollment cut-off based on the child’s month of birth. We use the following specification:

*First stage:*

$$YrsofSchooling_s = \alpha_0 + \alpha_1 MOB_s + \alpha_2 MOB_s^2 + \alpha_3 1[MOB_s \leq 6] + \mu_s \quad (4)$$

*Second stage:*

$$\begin{aligned} Correct_{ils} = & \beta_0 + \beta_1 Yrsof\widehat{Schooling}_s + \beta_2 PredictedDecline_l \\ & + \beta_3 Yrsof\widehat{Schooling}_s PredictedDecline_l + \beta_4 MOB_s + \beta_5 MOB_s^2 \\ & + \beta_6 MOB_s * PredictedDecline_l + \beta_7 MOB_s^2 * PredictedDecline_l + \epsilon_{ils} \end{aligned} \quad (5)$$

where  $YrsofSchooling_s$  captures the total years of schooling student  $s$  has received at the time of the exam.  $MOB_s$  is the student’s month of birth, which is the running variable in the regression discontinuity framework.  $1[MOB_s \leq 6]$  is an indicator that equals one if the student was born in the second half of the year (i.e. July to December, before the cut-off) and zero otherwise.<sup>45</sup>  $Correct_{ils}$  is a binary variable that captures whether student  $s$  correctly answered question item  $i$  appearing in location (decile)  $l$ .  $PredictedDecline_l$  is calculated in a parallel fashion to Eq. 2, where for each test subject, we take the difference between the average score in decile 1 minus the average score in decile  $l$ . We use students who are born in May through August as our “control group” to calculate Predicted Decline. The coefficient of interest is  $\beta_3$ , which captures the extent to which an additional year of schooling improves performance on the parts of the test where declines have set in.

### *Details of Instructional Quality Metrics*

The overall instructional quality metric is based on 12 items, all of which are considered to be markers of good quality instruction and have been shown to correlate with learning outcomes. These include dimensions such as classroom climate, feedback provided to students, time on task, use of higher-order thinking skills, etc. Brown and Andrabi (2021) describes the scoring and quality assurance process used in reviewing the classroom videos. Note that while all 66 schools are part of the same private school chain, there is substantial heterogeneity across them, with different schools serving different demographics and charging different levels of school fees. Other work has shown wealthier schools are associated with higher quality scores on the CLASS rubric (Cloney et al., 2016). However, this dataset does not contain direct information on socioeconomic status in this population.

---

<sup>45</sup>The sample is restricted to students born from July 2007 to June 2009.  $MOB$  begins at 1 for July 2007 goes up to 12 for June 2008, and then resets to 1 for July 2008 and goes up to 12 again for June 2009. The definition of  $1[MOB_s \leq 6]$  identifies a discrete jump at January in expected years of schooling. We use the 6 months before/after each January 1 cut-off to create non-overlapping samples of treatment and control students for the stacked RD across grades.