The Effects of Computer-Assisted Learning on Student Achievement in India
Pre-Analysis Plan

Table 1: Sample Construction

Sample Construction	Schools			Removed Schools		
	Total	**Treated**	**Control**	**Total**	**Treated**	**Control**
Sample 1: EL Analysis						
Initial sample	83	28	55	—	—	—
Remove: Schools without endline data	80	28	52	3	0	3
Final Sample 1	*80*	*28*	*52*	—	—	—
Sample 2: BL & EL Analysis						
Initial sample	83	28	55	—	—	—
Remove: Schools without endline data	80	28	52	3	0	3
Remove: Baseline cheating schools	74	24	50	6	4	2
Final Sample 2	*74*	*24*	*50*	—	—	—

Notes: Main samples constructed from 83 schools (28 treatment, 55 control) randomized to teatment intervention. Sample 1 focuses on endline (EL) analysis only, excluding schools without endline tests. Sample 2 enables baseline (BL) and endline (EL) analysis by additionally removing schools with suspected baseline cheating. Final sample sizes: Sample 1 (7,888 students), Sample 2 (5,535 students). Additional robustness samples excluding discontinued treatment schools are presented in Appendix Table 6.

Table 2: School-Level Balance Tests

Variable	Sample 1			Sample 2		
	Control Mean [SD]	Diff. (SE)	Diff. (SE)	Control Mean [SD]	Diff. (SE)	Diff. (SE)
District HDI	0.597	0.008	0.005	0.596	0.009	0.004
	[0.051]	(0.011)	(0.008)	[0.052]	(0.013)	(0.009)
All-Girls Schools	0.288	-0.003	0.009	0.280	0.012	0.009
		(0.107)	(0.106)		(0.114)	(0.114)
CBSE Curriculum	0.462	0.038	0.054	0.480	0.062	0.061
		(0.119)	(0.122)		(0.126)	(0.135)
School Size	94.519	11.659	11.624	71.040	11.585	11.458
	[38.652]	(8.010)	(8.166)	[29.599]	(7.529)	(7.684)
Missing Baseline (%)	23.810	1.748	1.933	—	—	—
	[10.363]	(2.527)	(2.540)			
Regression specifications						
District Group Fixed Effects			✓			✓
School Level Clustered SEs		✓	✓		✓	✓

Notes: School-level balance tests regressing each school characteristic on treatment assignment with school-level clustering. Each cell shows control group means with standard deviations in square brackets for continuous variables and treatment-control differences with standard errors in parentheses. District Group Fixed Effects control for geographic clustering in treatment assignment. District HDI is Human Development Index (0-1 scale). All-Girls Schools is proportion of all-girls schools versus all-boys schools (binary variable). CBSE Curriculum is proportion using CBSE (Central Board of Secondary Education) versus UP (Uttar Pradesh) Board curriculum (binary variable). School size is number of students per school. Missing baseline applies only to Sample 1 (Sample 2 requires complete baseline data). Sample 1 size: 80 schools (28 treatment, 52 control). Sample 2 size: 74 schools (24 treatment, 50 control). No significance stars appear as all differences are non-significant at conventional levels, indicating successful randomization. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 3: Student-Level Baseline Balance Tests

	Sample 1		Sample 2	
Control Mean	0.000	0.000	0.000	0.000
[SD]	[1.000]	[1.000]	[1.000]	[1.000]
Diff.	0.089	0.098	0.013	0.009
(SE)	(0.180)	(0.126)	(0.153)	(0.101)
Regression specifications				
District Group Fixed Effects		✓		✓
School Level Clustered SEs	✓	✓	✓	✓

Notes: Student-level baseline balance tests regressing standardized baseline test scores on treatment assignment. Diff. shows treatment-control differences with school-clustered standard errors in parentheses. Control group means are set to 0.000 and standard deviations (in square brackets) to 1.000 due to standardization of baseline scores. Sample 1 includes 5,980 students with available baseline data from the total 7,888 students with complete endline data. Sample 2 includes 5,535 students with complete baseline and endline data. The analysis focuses only on students with baseline data to test randomization effectiveness. District Group Fixed Effects control for geographic clustering in treatment assignment. No significance stars appear as all differences are non-significant at conventional levels ($p > 0.10$), indicating successful randomization. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 4: Balance of Baseline Exclusions

	Sample 2	
Diff.	-0.007	0.009
(SE)	(0.027)	(0.022)
Regression specifications		
District Group Fixed Effects		✓
School Level Clustered SEs	✓	✓

Notes: Balance test regressing whether students were excluded for missing baseline scores on treatment assignment to test if baseline exclusions are balanced between treatment and control groups. Diff. shows treatment-control difference in exclusion rates (treatment minus control) with school-clustered standard errors in parentheses. Among 7,289 students with complete endline scores eligible for Sample 2, 1,754 students were excluded for missing baseline scores: 1,141 control students (24.3%) and 613 treatment students (23.6%). This exclusion process resulted in Sample 2's final size of 5,535 students with complete baseline and endline data. District Group Fixed Effects control for geographic clustering in treatment assignment. No significance stars appear as differences are non-significant ($p > 0.10$), indicating that baseline exclusions are balanced. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 5: Balance of Excluded Schools

	Sample 1		Sample 2	
Variable	**Diff. (SE)**	**Diff. (SE)**	**Diff. (SE)**	**Diff. (SE)**
District HDI	0.002	0.001	0.003	-0.001
	(0.016)	(0.011)	(0.017)	(0.012)
All-Girls Schools	0.021	0.019	0.035	0.019
	(0.152)	(0.149)	(0.156)	(0.154)
CBSE Curriculum	-0.025	-0.015	-0.002	-0.008
	(0.167)	(0.172)	(0.172)	(0.182)
School Size	-1.067	-1.124	0.904	2.364
	(12.828)	(13.089)	(13.394)	(14.045)
Regression specifications				
District Group Fixed Effects		✓		✓
School Level Clustered SEs	✓	✓	✓	✓

Notes: Tests whether excluding schools changes treatment-control balance by estimating separate regressions of school characteristics on treatment assignment using the full sample (before exclusions) and final sample (after exclusions), then conducting t-tests comparing the treatment coefficients between the two models. Each cell shows the change in treatment-control difference with standard errors in parentheses calculated from the two regression standard errors. Sample 1 excludes 3 schools from original 83 schools. Sample 2 excludes 9 schools from original 83 schools. District HDI is Human Development Index (0-1 scale). All-Girls Schools is proportion of all-girls schools versus all-boys schools. CBSE Curriculum is proportion using CBSE (Central Board of Secondary Education) versus UP (Uttar Pradesh) Board curriculum. School size is number of students per school. District Group Fixed Effects control for geographic clustering in treatment assignment. No significance stars appear as all changes are non-significant ($p > 0.10$), indicating that school exclusions preserve randomization balance. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.
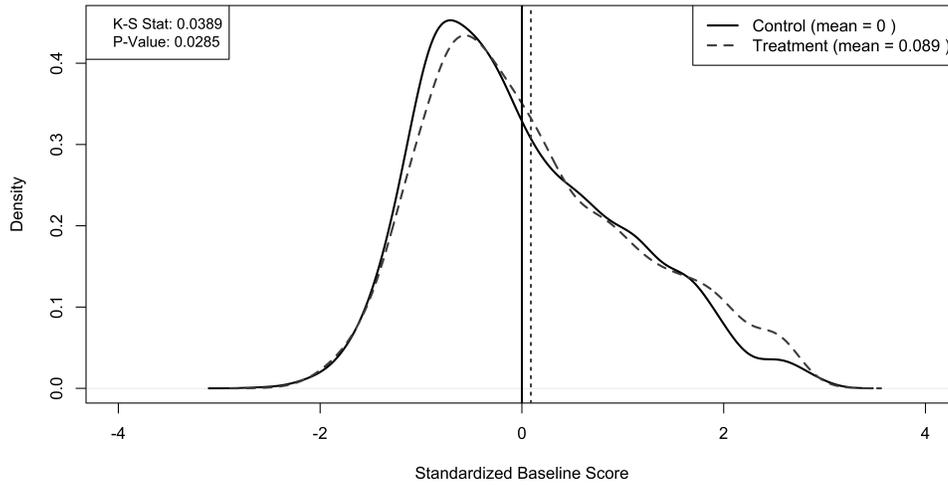
Figure 1: Sample 1

Notes: Kernel density plots comparing standardized baseline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means (control mean = 0, treatment mean = 0.089). Sample 1 includes 5,980 students with available baseline data. Kolmogorov-Smirnov test statistic of 0.0389 with p-value of 0.0285 is significant at the 5% level.
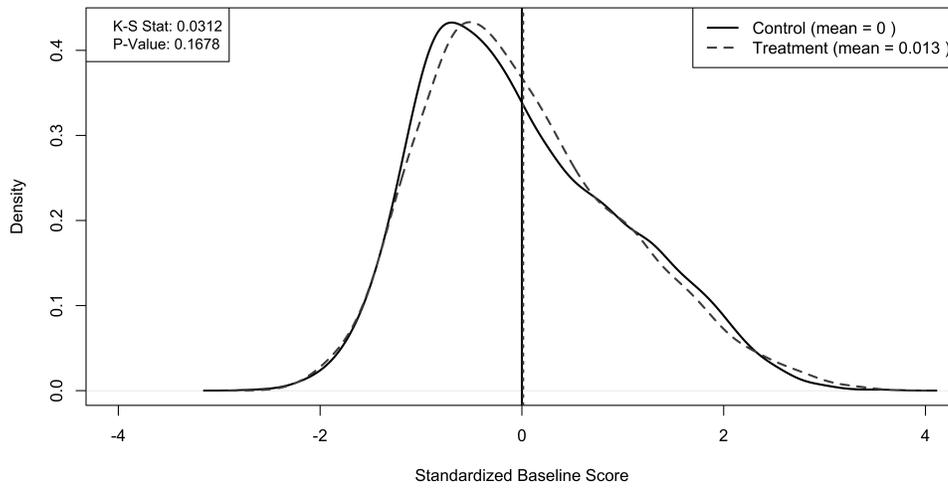

Figure 2: Sample 2

Notes: Kernel density plots comparing standardized baseline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means (control mean = 0, treatment mean = 0.013). Sample 2 includes 5,535 students with available baseline data. Kolmogorov-Smirnov test statistic of 0.0312 with p-value of 0.1678 is not significant at conventional levels.

APPENDIX

Table 6: Sample Construction - Robustness Analysis

	Schools			Removed Schools		
Sample Construction	**Total**	**Treated**	**Control**	**Total**	**Treated**	**Control**
Sample 3: EL Analysis						
Initial sample	83	28	55	—	—	—
Remove: Schools without endline data	80	28	52	3	0	3
Remove: Discontinued treatment schools	75	23	52	5	5	0
Final Sample 3	*75*	*23*	*52*	—	—	—
Sample 4: BL & EL Analysis						
Initial sample	83	28	55	—	—	—
Remove: Schools without endline data	80	28	52	3	0	3
Remove: Discontinued treatment schools	75	23	52	5	5	0
Remove: Baseline cheating schools	73	23	50	2	0	2
Final Sample 4	*73*	*23*	*50*	—	—	—

Notes: Robustness analysis samples that additionally exclude discontinued treatment schools. Discontinued treatment schools are those where the treatment could not continue due to technical barriers such as lack of digital devices, internet unavailability, or insufficient support. Sample 3 excludes schools without endline data and discontinued treatment schools for endline (EL) analysis. Sample 4 excludes schools without endline data, discontinued treatment schools, and schools with suspected baseline cheating to enable baseline (BL) and endline (EL) analysis. Final sample sizes: Sample 3 (7,399 students), Sample 4 (5,453 students).

Table 7: School-Level Balance Tests - Robustness Analysis

	Sample 3			Sample 4		
Variable	**Control Mean [SD]**	**Diff. (SE)**	**Diff. (SE)**	**Control Mean [SD]**	**Diff. (SE)**	**Diff. (SE)**
District HDI	0.597	0.008	0.003	0.596	0.009	0.003
	[0.051]	(0.013)	(0.010)	[0.052]	(0.013)	(0.010)
All-Girls Schools	0.288	-0.028	-0.014	0.280	-0.019	-0.014
		(0.113)	(0.116)		(0.113)	(0.116)
CBSE Curriculum	0.462	0.060	0.041	0.480	0.042	0.041
		(0.127)	(0.139)		(0.128)	(0.139)
School Size	94.519	13.481	16.492*	71.040	11.612	12.294
	[38.652]	(8.527)	(9.031)	[29.599]	(7.757)	(7.958)
Missing Baseline (%)	23.810	0.580	1.287	—	—	—
	[10.363]	(2.742)	(2.914)			
Regression specifications						
District Group Fixed Effects			✓			✓
School Level Clustered SEs		✓	✓		✓	✓

Notes: School-level balance tests regressing each school characteristic on treatment assignment with school-level clustering. Each cell shows control group means with standard deviations in square brackets for continuous variables and treatment-control differences with standard errors in parentheses. District Group Fixed Effects control for geographic clustering in treatment assignment. District HDI is Human Development Index (0-1 scale). All-Girls Schools is proportion of all-girls schools versus all-boys schools (binary variable). CBSE Curriculum is proportion using CBSE (Central Board of Secondary Education) versus UP (Uttar Pradesh) Board curriculum (binary variable). School size is number of students per school. Missing baseline applies only to Sample 3 (Sample 4 requires complete baseline data). Sample 3 size: 75 schools (23 treatment, 52 control). Sample 4 size: 73 schools (23 treatment, 50 control). One coefficient shows significance at the 10% level (School Size in Sample 3 with district fixed effects) with all others showing no significance, indicating successful randomization. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 8: Student-Level Baseline Balance Tests - Robustness Analysis

	Sample 3		Sample 4	
Control Mean	0.000	0.000	0.000	0.000
[SD]	[1.000]	[1.000]	[1.000]	[1.000]
Diff.	-0.067	0.014	0.030	0.017
(SE)	(0.158)	(0.096)	(0.156)	(0.105)
Regression specifications				
District Group Fixed Effects		✓		✓
School Level Clustered SEs	✓	✓	✓	✓

Notes: Student-level baseline balance tests regressing standardized baseline test scores on treatment assignment for robustness samples excluding discontinued treatment schools. Diff. shows treatment-control differences with school-clustered standard errors in parentheses. Control group means are set to 0.000 and standard deviations (in square brackets) to 1.000 due to standardization of baseline scores. Sample 3 includes 5,633 students with available baseline data from the total 7,399 students with complete endline data. Sample 4 includes 5,453 students with complete baseline and endline data. The analysis focuses only on students with baseline data to test randomization effectiveness. District Group Fixed Effects control for geographic clustering in treatment assignment. No significance stars appear as all differences are non-significant at conventional levels ($p > 0.10$), indicating successful randomization. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 9: Balance of Baseline Exclusions - Robustness Analysis

	Sample 4	
Diff.	-0.008	0.009
(SE)	(0.028)	(0.023)
Regression specifications		
District Group Fixed Effects		✓
School Level Clustered SEs	✓	✓

Notes: Balance test regressing whether students were excluded for missing baseline scores on treatment assignment to test if baseline exclusions are balanced between treatment and control groups for robustness samples excluding discontinued treatment schools. Diff. shows treatment-control difference in exclusion rates (treatment minus control) with school-clustered standard errors in parentheses. Among 7,177 students with complete endline scores eligible for Sample 4, 1,724 students were excluded for missing baseline scores: 1,141 control students (24.3%) and 583 treatment students (23.5%). This exclusion process resulted in Sample 4's final size of 5,453 students with complete baseline and endline data. District Group Fixed Effects control for geographic clustering in treatment assignment. No significance stars appear as differences are non-significant ($p > 0.10$), indicating that baseline exclusions are balanced. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 10: Balance of Excluded Schools - Robustness Analysis

Variable	Sample 3		Sample 4	
	Diff. (SE)	Diff. (SE)	Diff. (SE)	Diff. (SE)
District HDI	0.002	-0.002	0.002	-0.002
	(0.017)	(0.012)	(0.017)	(0.012)
All-Girls Schools	-0.004	-0.003	0.004	-0.003
	(0.155)	(0.156)	(0.156)	(0.156)
CBSE Curriculum	-0.003	-0.028	-0.022	-0.028
	(0.173)	(0.184)	(0.173)	(0.185)
School Size	0.163	5.522	1.741	5.522
	(13.457)	(14.153)	(13.543)	(14.186)
Regression specifications				
District Group Fixed Effects		✓		✓
School Level Clustered SEs	✓	✓	✓	✓

Notes: Tests whether excluding schools changes treatment-control balance for robustness samples excluding discontinued treatment schools. We estimate separate regressions of school characteristics on treatment assignment using the full sample (before exclusions) and final sample (after exclusions), then conduct t-tests comparing the treatment coefficients between the two models. Each cell shows the change in treatment-control difference with standard errors in parentheses calculated from the two regression standard errors. Sample 3 excludes 8 schools. Sample 4 excludes 10 schools. District HDI is Human Development Index (0-1 scale). All-Girls Schools is proportion of all-girls schools versus all-boys schools. CBSE Curriculum is proportion using CBSE (Central Board of Secondary Education) versus UP (Uttar Pradesh) Board curriculum. School size is number of students per school. District Group Fixed Effects control for geographic clustering in treatment assignment. No significance stars appear as all changes are non-significant ($p > 0.10$), indicating that school exclusions preserve randomization balance. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.
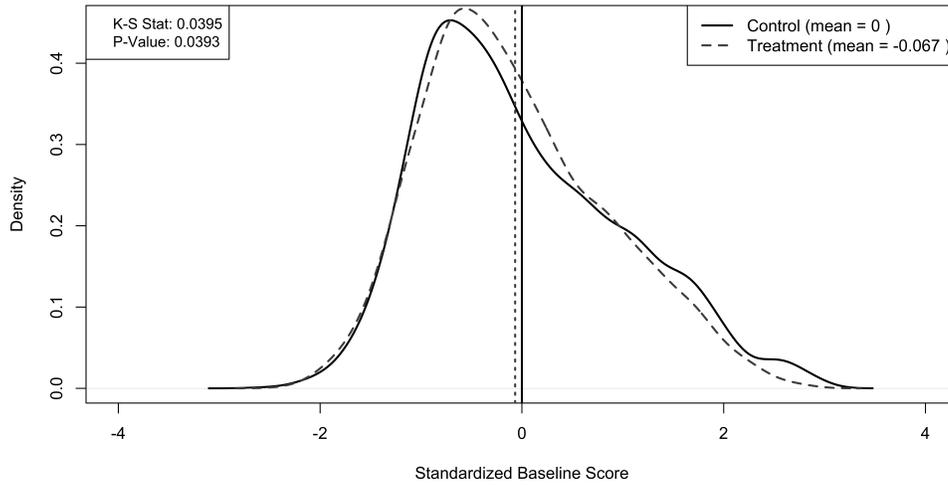
Figure 3: Sample 3

Notes: Kernel density plots comparing standardized baseline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means (control mean = 0, treatment mean = -0.067). Sample 3 includes 5,633 students with available baseline data. Kolmogorov-Smirnov test statistic of 0.0395 with p-value of 0.0393 is significant at the 5% level.

Figure 4: Sample 4

Notes: Kernel density plots comparing standardized baseline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means (control mean = 0, treatment mean = 0.03). Sample 4 includes 5,453 students with available baseline data. Kolmogorov-Smirnov test statistic of 0.0338 with p-value of 0.1174 is not significant at conventional levels.
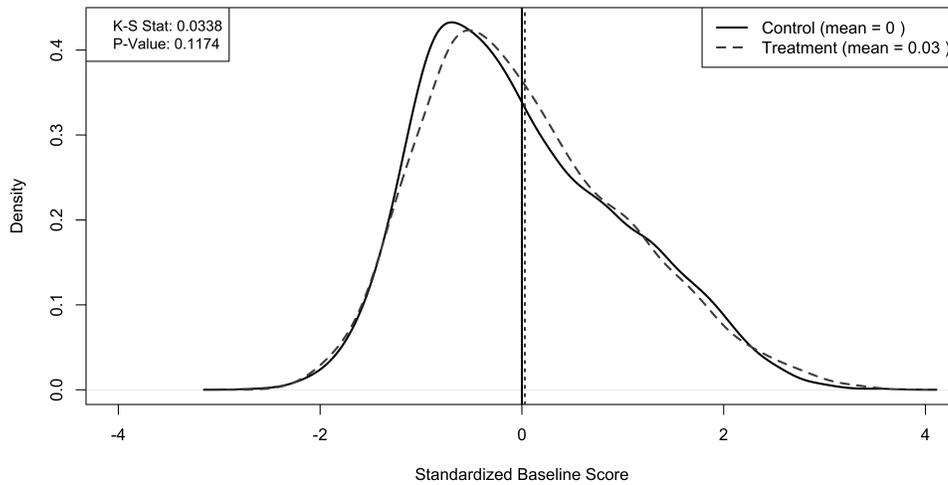
RESULTS

Table 11: Intent-to-Treat Effects on Endline Test Scores

	Sample 1			Sample 2		
	ITT (1)	ITT (2)	ITT (3)	ITT (4)	ITT (5)	ITT (6)
Diff.	-	-	-	-	-	-
(SE)	(-)	(-)	(-)	(-)	(-)	(-)
Control Mean	-	-	-	-	-	-
[SD]	[-]	[-]	[-]	[-]	[-]	[-]
Observations	-	-	-	-	-	-
Regression specifications						
Std. Baseline Scores			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Intent-to-treat effects on standardized endline test scores. Each column shows results from separate regressions comparing all students in treatment vs control schools. Diff. shows treatment-control differences with school-clustered standard errors in parentheses. Control group means and standard deviations in square brackets shown for standardized endline scores. Sample 1 excludes schools without endline tests (7,888 students from 80 schools). Sample 2 additionally excludes schools with suspected baseline cheating (5,535 students from 74 schools). Std. Baseline Scores include standardized baseline test scores as controls. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 12: Treatment-on-Treated Effects on Endline Test Scores

	Sample 1			Sample 2		
	TOT (1)	TOT (2)	TOT (3)	TOT (4)	TOT (5)	TOT (6)
Diff.	-	-	-	-	-	-
(SE)	(-)	(-)	(-)	(-)	(-)	(-)
Control Mean	-	-	-	-	-	-
[SD]	[-]	[-]	[-]	[-]	[-]	[-]
Observations	-	-	-	-	-	-
Regression specifications						
Std. Baseline Scores			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Treatment-on-treated effects on standardized endline test scores using treatment assignment as instrument for actual treatment receipt. Each column shows results from separate instrumental variable regressions. Diff. shows treatment-control differences with school-clustered standard errors in parentheses. Control group means and standard deviations in square brackets shown for standardized endline scores. Sample 1 excludes schools without endline tests (7,888 students from 80 schools). Sample 2 additionally excludes schools with suspected baseline cheating (5,535 students from 74 schools). Std. Baseline Scores include standardized baseline test scores as controls. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 13: Heterogeneous Intent-to-Treat Effects on Endline Test Scores

	Sample 1			Sample 2		
	ITT (1)	ITT (2)	ITT (3)	ITT (4)	ITT (5)	ITT (6)
Panel A: Gender						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × All-Girls	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel B: HDI						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Above Median HDI	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel C: Curriculum						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × CBSE	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel D: Baseline Performance						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Low Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Medium Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Observations	-	-	-	-	-	-
Regression specifications						
Baseline Controls			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Heterogeneous intent-to-treat effects on standardized endline test scores. Each panel shows results from separate regressions including the specified interaction terms. Treatment Effect shows the main treatment effect for the reference group. Interaction terms show the additional treatment effect for the specified subgroup. Total treatment effect for interaction group equals Treatment Effect + Interaction coefficient. Reference groups: All-Boys schools (Panel A), Below Median HDI (Panel B), UP Board curriculum (Panel C), High Tercile (Panel D). Students classified into terciles based on baseline test scores: Low Tercile = bottom tercile (0-33rd percentile), Medium Tercile = middle tercile (34th-67th percentile), High Tercile = top tercile (68th-100th percentile). Sample 1 excludes schools without endline tests (7,888 students from 80 schools). Sample 2 additionally excludes schools with suspected baseline cheating (5,535 students from 74 schools). Baseline controls include standardized baseline test scores. Significance levels: *** p ≤ 0.01, ** p ≤ 0.05, * p ≤ 0.10.

Table 14: Heterogeneous Treatment-on-Treated Effects on Endline Test Scores

	Sample 1			Sample 2		
	TOT (1)	TOT (2)	TOT (3)	TOT (4)	TOT (5)	TOT (6)
Panel A: Gender						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × All-Girls	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel B: HDI						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Above Median HDI	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel C: Curriculum						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × CBSE	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel D: Baseline Performance						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Low Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Medium Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Observations	-	-	-	-	-	-
Regression specifications						
Baseline Controls			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Heterogeneous treatment-on-treated effects on standardized endline test scores. Each panel shows results from separate regressions including the specified interaction terms. Treatment Effect shows the main treatment effect for the reference group. Interaction terms show the additional treatment effect for the specified subgroup. Total treatment effect for interaction group equals Treatment Effect + Interaction coefficient. Reference groups: All-Boys schools (Panel A), Below Median HDI (Panel B), UP Board curriculum (Panel C), High Tercile (Panel D). Students classified into terciles based on baseline test scores: Low Tercile = bottom tercile (0-33rd percentile), Medium Tercile = middle tercile (34th-67th percentile), High Tercile = top tercile (68th-100th percentile). Sample 1 excludes schools without endline tests (7,888 students from 80 schools). Sample 2 additionally excludes schools with suspected baseline cheating (5,535 students from 74 schools). Baseline controls include standardized baseline test scores. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Figure 5: Sample 1

Notes: Kernel density plots comparing standardized endline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means. Sample 1 includes 7,888 students with complete endline data.

density_plot_sample_2_endline.png

Figure 6: Sample 2

Notes: Kernel density plots comparing standardized endline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means. Sample 2 includes 5,535 students with complete endline data.

RESULTS APPENDIX

Table 15: Intent-to-Treat Effects on Endline Test Scores
Robustness Analysis

	Sample 3			Sample 4		
	ITT (1)	ITT (2)	ITT (3)	ITT (4)	ITT (5)	ITT (6)
Diff.	-	-	-	-	-	-
(SE)	(-)	(-)	(-)	(-)	(-)	(-)
Control Mean	-	-	-	-	-	-
[SD]	[-]	[-]	[-]	[-]	[-]	[-]
Observations	-	-	-	-	-	-
Regression specifications						
Std. Baseline Scores			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Intent-to-treat effects on standardized endline test scores for robustness samples excluding discontinued treatment schools. Each column shows results from separate regressions comparing all students in treatment vs control schools. Diff. shows treatment-control differences with school-clustered standard errors in parentheses. Control group means and standard deviations in square brackets shown for standardized endline scores. Sample 3 excludes schools without endline tests and discontinued treatment schools (7,399 students from 75 schools). Sample 4 additionally excludes schools with suspected baseline cheating (5,453 students from 73 schools). Std. Baseline Scores include standardized baseline test scores as controls. Significance levels: *** p ≤ 0.01, ** p ≤ 0.05, * p ≤ 0.10.

Table 16: Treatment-on-Treated Effects on Endline Test Scores
Robustness Analysis

	Sample 3			Sample 4		
	TOT (1)	TOT (2)	TOT (3)	TOT (4)	TOT (5)	TOT (6)
Diff.	-	-	-	-	-	-
(SE)	(-)	(-)	(-)	(-)	(-)	(-)
Control Mean	-	-	-	-	-	-
[SD]	[-]	[-]	[-]	[-]	[-]	[-]
Observations	-	-	-	-	-	-
Regression specifications						
Std. Baseline Scores			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Treatment-on-treated effects on standardized endline test scores for robustness samples excluding discontinued treatment schools using treatment assignment as instrument for actual treatment receipt. Each column shows results from separate instrumental variable regressions. Diff. shows treatment-control differences with school-clustered standard errors in parentheses. Control group means and standard deviations in square brackets shown for standardized endline scores. Sample 3 excludes schools without endline tests and discontinued treatment schools (7,399 students from 75 schools). Sample 4 additionally excludes schools with suspected baseline cheating (5,453 students from 73 schools). Std. Baseline Scores include standardized baseline test scores as controls. Significance levels: *** p ≤ 0.01, ** p ≤ 0.05, * p ≤ 0.10.

Table 17: Heterogeneous Intent-to-Treat Effects on Endline Test Scores
Robustness Analysis

	Sample 3			Sample 4		
	ITT (1)	ITT (2)	ITT (3)	ITT (4)	ITT (5)	ITT (6)
Panel A: Gender						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × All-Girls	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel B: HDI						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Above Median HDI	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel C: Curriculum						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × CBSE	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel D: Baseline Performance						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Low Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Medium Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Observations	-	-	-	-	-	-
Regression specifications						
Baseline Controls			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Heterogeneous intent-to-treat effects on standardized endline test scores for robustness samples excluding discontinued treatment schools. Each panel shows results from separate regressions including the specified interaction terms. Treatment Effect shows the main treatment effect for the reference group. Interaction terms show the additional treatment effect for the specified subgroup. Total treatment effect for interaction group equals Treatment Effect + Interaction coefficient. Reference groups: All-Boys schools (Panel A), Below Median HDI (Panel B), UP Board curriculum (Panel C), High Tercile (Panel D). Students classified into terciles based on baseline test scores: Low Tercile = bottom tercile (0-33rd percentile), Medium Tercile = middle tercile (34th-67th percentile), High Tercile = top tercile (68th-100th percentile). Sample 3 excludes schools without endline tests and discontinued treatment schools (7,399 students from 75 schools). Sample 4 additionally excludes schools with suspected baseline cheating (5,453 students from 73 schools). Baseline controls include standardized baseline test scores. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table 18: Heterogeneous Treatment-on-Treated Effects on Endline Test Scores
Robustness Analysis

	Sample 3			Sample 4		
	TOT (1)	TOT (2)	TOT (3)	TOT (4)	TOT (5)	TOT (6)
Panel A: Gender						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × All-Girls	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel B: HDI						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Above Median HDI	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel C: Curriculum						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × CBSE	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Panel D: Baseline Performance						
Treatment Effect	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Low Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Treatment × Medium Tercile	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
Observations	-	-	-	-	-	-
Regression specifications						
Baseline Controls			✓			✓
District Fixed Effects		✓	✓		✓	✓
School Level Clustered SEs	✓	✓	✓	✓	✓	✓

Notes: Heterogeneous treatment-on-treated effects on standardized endline test scores for robustness samples excluding discontinued treatment schools. Each panel shows results from separate regressions including the specified interaction terms. Treatment Effect shows the main treatment effect for the reference group. Interaction terms show the additional treatment effect for the specified subgroup. Total treatment effect for interaction group equals Treatment Effect + Interaction coefficient. Reference groups: All-Boys schools (Panel A), Below Median HDI (Panel B), UP Board curriculum (Panel C), High Tercile (Panel D). Students classified into terciles based on baseline test scores: Low Tercile = bottom tercile (0-33rd percentile), Medium Tercile = middle tercile (34th-67th percentile), High Tercile = top tercile (68th-100th percentile). Sample 3 excludes schools without endline tests and discontinued treatment schools (7,399 students from 75 schools). Sample 4 additionally excludes schools with suspected baseline cheating (5,453 students from 73 schools). Baseline controls include standardized baseline test scores. Significance levels: *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Figure 7: Sample 3

Notes: Kernel density plots comparing standardized endline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means. Sample 3 includes 7,399 students with complete endline data.

density_plot_sample_4_endline.png

Figure 8: Sample 4

Notes: Kernel density plots comparing standardized endline test score distributions between treatment and control groups. Solid line shows control group distribution, dashed line shows treatment group distribution. Vertical lines indicate group means. Sample 4 includes 5,453 students with complete endline data.