The impact of river water quality on children's education: evidence from 39 districts in India

Md Ohiul Islam[1*] and Moumita Ghorai[2]

[1]Economics Department, University of Nevada at Reno, Reno, NV, USA and [2]United Nations Development Programme, New York, NY, USA

[*]Corresponding author. E-mail: mdohiul.islam@wmich.edu; oislam@unr.edu

Online Appendix

Table A1. Districts with unsafe levels of pollution

District	1[*FCOLI* > *limit*]	1[*NIT* > *limit*]	Obs.
Bihar - Purbi Champaran	1	1	49
Bihar – Madhubani	0	1	51
Bihar – Supaul	0	1	17
Bihar - Muzaffarpur	0	1	56
Bihar – Siwan	0	1	32
Bihar – Saran	0	1	7
Bihar – Bhagalpur	1	1	43
Bihar – Patna	0	1	10
Bihar – Buxar	1	1	11
Bihar – Rohtas	0	1	58
Bihar – Gaya	0	1	13
Uttar Pradesh - Muzaffarnagar	1	1	33
Uttar Pradesh - Meerut	1	0	11
Uttar Pradesh - Ghaziabad	1	0	28
Uttar Pradesh - Mathura	1	1	44
Uttar Pradesh - Budaun	0	1	17
Uttar Pradesh - Lucknow	1	0	12
Uttar Pradesh - Kannauj	1	0	23
Uttar Pradesh - Kanpur Nagar	1	0	47
Uttar Pradesh - Jhansi	1	1	16
Uttar Pradesh - Prayagraj	1	1	67
Uttar Pradesh - Faizabad	1	1	51
Uttar Pradesh - Gorakhpur	0	0	19
Uttar Pradesh - Deoria	0	1	10
Uttar Pradesh - Varansi	1	1	44
Uttarakhand - Dehradun	1	1	11
West Bengal - Darjiling	1	1	22
West Bengal - Jalpaiguri	1	1	33
West Bengal - Murshidabad	1	0	38
West Bengal - Birbhum	1	0	31
West Bengal - Barddhaman	1	0	63
West Bengal - Nadia	1	1	30
West Bengal - North 24 Parganas	1	1	26
West Bengal - Hugli	1	0	19
West Bengal - Bankura	1	1	6
West Bengal - Haora	1	0	15
West Bengal - Kolkata	1	1	36
West Bengal - South 24 Parganas	1	1	12
Jharkhand - Palamu	0	1	36
N			1147

Explanation for table A1

Columns 1 and 2 indicate whether a district listed in the table experiences, on average, unsafe levels of mean faecal coliform (FCOLI) and mean Nitrate-N + Nitrite-N (NIT), respectively. Districts experiencing unsafe levels of a pollutant are assigned a 1, and zero otherwise. Column 3 displays the number of observations for each district.

Table A2. Analytical sample means of key variables

Correlation matrix of water quality variables					
District-level means in rows and columns	FCOL	NIT	DO	BOD	pH
FCOL	1				
NIT	-0.0411	1			
DO	-0.2391	-0.0497	1		
BOD	0.1958	0.3061	-0.667	1	
pH	-0.0386	0.0801	0.2494	-0.1206	1

Binary indicators for water quality 1 = unsafe levels of district-mean reading	(1) FCOLI=1	(2) NIT=1	(3) DO=1	(4) BOD=1	(5) pH=1
Key variables					
Mean faecal coliform (Millions MPN/100 mL) †	1.61	0.38	4.99	2.66	1.19
Mean Nitrate-N/Nitrite-N (mg/l)	1.20	2.77	0.64	1.58	0.92
Mean BOD (mg/l)	5.79	6.21	11.57	8.15	4.96
Mean DO (mg/l)	6.56	6.54	3.91	5.95	6.94
Mean pH	7.65	7.58	7.55	7.74	7.69
Reading score	0.16	0.36	0.14	0.11	0.14
Maths score	0.21	0.35	0.21	0.11	0.19
Writing score	0.18	0.29	0.10	0.09	0.15
Age	9.49	9.59	9.47	9.55	9.48
Sex - Male	0.51	0.52	0.51	0.55	0.51
1[Majority religious group]	0.50	0.38	0.53	0.57	0.52
Anthropometry - height	127.15	126.41	126.87	127.16	127.01
Anthropometry - weight	25.38	25.23	25.44	25.54	25.34
1[*HH per capita expenditure* $\leq$ *25th ptile*]	0.25	0.22	0.14	0.216	0.26
1[*HH per capita expenditure* $\leq$ *50th ptile*]	0.49	0.47	0.40	0.44	0.51
1[*HH per capita expenditure* $\leq$ *75th ptile*]	0.73	0.75	0.67	0.70	0.76
School distance (km)	1.66	1.55	1.69	1.66	1.57
School hours/week	29.38	28.12	29.28	30.98	30.27
Private tuition hours/week	9.27	9.90	7.88	7.54	8.67
Books uniform cost	857.11	764.79	1002.63	983.04	822.01
Short-term morbidity (days)	1.13	1.08	0.83	1.19	1.08
1 = HH purifies water before drinking	0.11	0.08	0.09	0.08	0.09
1 = HH indoor piped drinking water	0.15	0.13	0.15	0.16	0.12
1 = HH has water storage vessels	0.79	0.84	0.73	0.78	0.70
1 = HH always wash hands after defaecation	0.72	0.69	0.75	0.77	0.72
Number of Observations	821	306	246	492	1111

Notes: Mean faecal coliform (MPN/100 ml), reported in millions. Pollution indicators. FCOLI = 1, NIT = 1, DO = 1, BOD = 1, pH = 1 mean unsafe levels of district-mean faecal coliform, Nitrate-N + Nitrite-N, dissolved oxygen, biochemical oxygen demand respectively.

Explanation for table A2

The first part of table A2 presents a correlation matrix for district-level mean water quality variables. Interestingly, certain pairs of district-level means, such as $FCOLI$ and NIT , display a negative correlation. In table A2, the column means are derived from samples where water quality indicator readings fell below safety and/or recommended levels. For example, the variable age corresponds to the mean $age|FCOLI = 1$ in column 1. It is observed that mean test scores are higher when $NIT = 1$ (column 2) as compared to when $FCOLI = 1$ (column 2). Generally, the variable means show minimal variation across the columns. Notably, the rate of water purification in households is 0.75 when the BOD exceeds its safe threshold, indicating that 75% of households treat their water before consumption. This percentage significantly surpasses that of households in the sample that purify their water when mean faecal coliform, mean Nitrate-N + Nitrite-N, and pH levels are not within safe limits.

Table A3. Analytical sample means of additional control variables

Variable	(1) 1[$\overline{FCOLI}$ > *limit*]	(2) 1[$\overline{NIT}$ > *limit*]	(3) River	(4) All
Short-term morbidity (STM)				
1[*Insurance Coverage for last* ***STM***]	0.12	0.12	0.17	0.09
STM: Total cost***	47.12	46.08	58.80	49.78
STM: Days ill in last 30 days	1.65	1.73	1.70	1.53
STM: Days with fever in last 30 days	0.26	0.26	0.29	0.25
STM: Days with Cough in last 30 days	0.22	0.22	0.23	0.20
STM: Additional costs ◊	26.57	27.01	34.97	30.80
STM: Additional costs(2) ◊◊	2.75	1.99	2.21	2.69
Water supply				
1[*Water supply adequate*]	0.95	0.98	0.98	0.96
Use of technology in the HH				
1[*HH respondent has mobile phone*]	0.36	0.28	0.40	0.32
1[*HH uses computer*]	0.05	0.04	0.07	0.05
Schooling and teaching-related				
If child gets free uniform (Yes = 1)	0.29	0.33	0.26	0.31
Years of education completed: None	0.079	0.084	0.081	0.086
Years of education completed: 1-4	0.73	0.73	0.73	0.73
Years of education completed: primary or 5	0.11	0.11	0.12	0.12
Years of Education completed: 6-9	0.069	0.059	0.069	0.063
Scholarship amount	76.98	75.60	76.63	75.76
School fees (in thousand Rupees)	1.6	1.4	1.8	1.4
Child's teacher **(CT)**				
1[***CT*** *is fair to him/her*]	0.10	0.07	0.11	0.13
1[***CT*** *is biased*]	0.04	0.06	0.04	0.06
1[***CT*** *is local*]	0.44	0.52	0.51	0.50
1[***CT*** *is female*]	0.36	0.37	0.39	0.37
Parents attended PTA meeting (*Yes* = 1)	0.41	0.36	0.45	0.37
School admission was difficult (*Yes* = 1)	0.28	0.32	0.26	0.31
Short-term morbidity ◊◊◊	1.69	0 8	1.70	1.55
Urban = 1 and Rural = 0	0.36	0.31	0.44	0.36
Number of Observations	821	306	576	1147

Notes: ***Mean short-term morbidity total cost for inpatient/outpatient (doctor and hospital). FCOLI = Faecal coliform, NIT = Nitrate-N + Nitrite-N, and HH = Household.

◊ Mean short-term morbidity additional costs including medicines/tests/expenses which are not included in item (***).

◊◊ Mean short-term morbidity travel expenses.

◊◊◊ District-wise average number of days spent disabled due to short-term morbidity.

Explanation for table A3

Column 1 shows means for the sample of districts with unsafe levels of faecal coliform, column 2 shows means for the sample of districts with unsafe levels of Nitrate-N + Nitrite-N, column 3 shows means for the sample of districts along rivers, and column 4 shows means for the sample of all districts.

Table A4. Pratham's assessment framework

Assessment Framework	The assessment framework comprises three tests, each tailored to evaluate a distinct set of skills or knowledge areas. Unlike a uniform scoring system, these tests feature varying numbers of tasks, resulting in different maximum scores. Scores start at 0, representing no tasks completed, with the highest possible score corresponding to the total number of tasks within each test.
Test Outcomes	**Total points on the reading test score is 4**. The score is based on the child's ability to read a story. In the analytical sample, 12.45% of children cannot read the story (score 0), 11.88% can only recognise the letters (score 1), 17.96% can read some words (score 2), 16.27% can read paragraphs (score 3), and 41.44% can read the entire story (score 4). **The total writing test score is 2**: 27.44% cannot write at all (score 0), 29.14% write with one or two mistakes (score 1), and 43.42% produce error-free writing (score 2). **The total maths test score is 3**. It evaluates computational skills. 19.52% cannot recognise any numbers (score 0), 30.98% can recognise numbers (score 1), 24.47% can perform subtractions (score 2), and 25.04% can execute divisions correctly (score 3).

Note: The tests were conducted by Pratham (2021), a non-governmental organisation in India.

Table A5. Interaction between indoor piped water supply and exposure to faecal Coliform above safety level

Test	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score	(7) Reading Score	(8) Maths Score	(9) Writing Score
Female	-0.0255	-0.0920	0.0306	-0.0175	-0.0866	0.0204	-0.0261	-0.0568	0.0778
	(0.0377)	(0.0495)	(0.0440)	(0.0577)	(0.0823)	(0.0682)	(0.0409)	(0.0598)	(0.0513)
(1) Household has Indoor Piped Water	0.130	-0.328	0.894	0.0447	-0.267	0.818	0.216	-0.296	0.819
	(0.486)	(0.406)	(0.163)	(0.391)	(0.344)	(0.190)	(0.533)	(0.412)	(0.191)
$1[\overline{FCOLI} > limit]$	-0.244	-0.210	-0.189	-0.173	-0.150	-0.444	-0.482	-0.323	-0.0411
	(0.0837)	(0.105)	(0.148)	(0.160)	(0.173)	(0.192)	(0.122)	(0.142)	(0.184)
$(1) \times 1[\overline{FCOLI} > limit]$	-0.00651	0.389	-0.812	0.259	0.387	-0.803	-0.0373	0.546	-0.630
	(0.466)	(0.405)	(0.167)	(0.380)	(0.397)	(0.211)	(0.492)	(0.438)	(0.236)
$1[\overline{NIT} > limit]$	0.0529	-0.0347	-0.129	0.171	0.0441	-0.122	0.0254	-0.00812	-0.230
	(0.0732)	(0.0756)	(0.0835)	(0.231)	(0.129)	(0.215)	(0.0689)	(0.0886)	(0.107)
Observations	1147	1147	1147	576	576	576	738	738	738
Samples	All	All	All	River	River	River	Trib.	Trib.	Trib.

Notes: Robust standard errors clustered at district level in parentheses. FCOLI = Faecal Coliform and NIT = Nitrate-N + Nitrite-N. The regression specifications corresponding the results in columns 1 to 9 include all explanatory variables used for results in tables 2, 3, and 4 with additional controls for (1) whether respondent child receives scholarship for education, (2) year-round water availability (1=adequate, 0=inadequate), (3) drinking water storage vessel (1=the household has storage vessel, 0=none), (4) Whether the HH boils water to purify water (1=does, 0=does not), (5) frequency of washing hands after defaecation, (6) completed years of education, (7) Binary: Whether household has mobile phones, (8) Binary: Whether household has computer, (9) Short-term morbidity controls - the number of days the child was disabled in the last thirty days, the number of days the child showed certain symptoms like fever and coughing, and the amount of medical cost due to the short-term morbidity in the last thirty days, (10) school fees, whether the child gets free uniform (binary), costs of books, (11) Whether the household head considers the child's class teacher to be fair, parents' PTA participation (binary), child's teacher's gender, whether the child's admission to school was difficulty, frequency of child's teacher being absent at school. Lastly, district-level average short-term morbidity, state fixed effects, and survey month controls are also included to estimate each result.

Table A6. Impact of water pollutants - differences between genders

All districts	(1) Score Reading	(2) Score Maths	(3) Score Writing	(4) Score Reading	(5) Score Maths	(6) Score Writing
$1[\overline{FCOLI} > limit]$	-0.0726	-0.284	-0.112	-0.128	-0.0508	-0.110
	(0.104)	(0.176)	(0.116)	(0.106)	(0.119)	(0.159)
$1[\overline{NIT} > limit]$	-0.192	-0.174	-0.0387	-0.0833	0.0943	0.116
	(0.0867)	(0.196)	(0.0981)	(0.141)	(0.144)	(0.155)
$1[\overline{D.O.} < threshold]$	-0.114	0.00571	-0.0914	0.0379	-0.0609	-0.180
	(0.0983)	(0.155)	(0.100)	(0.109)	(0.137)	(0.118)
Mean pH	-0.102	-0.190	-0.0373	-0.199	-0.526	-0.242
	(0.123)	(0.245)	(0.0806)	(0.150)	(0.172)	(0.146)
N	592	592	592	555	555	555
Gender	Male	Male	Male	Female	Female	Female
Overall R^2	0.30	0.27	0.27	0.19	0.26	0.29

Districts near rivers	(7) Score Reading	(8) Score Maths	(9) Score Writing	(10) Score Reading	(11) Score Maths	(12) Score Writing
$1[\overline{FCOLI} > limit]$	0.0329	-0.183	-0.464	-0.534	-0.582	-0.369
	(0.164)	(0.165)	(0.216)	(0.105)	(0.153)	(0.182)
$1[\overline{NIT} > limit]$	0.0214	0.151	-0.0944	-0.231	-0.0863	0.153
	(0.149)	(0.113)	(0.203)	(0.237)	(0.162)	(0.202)
$1[\overline{D.O.} < threshold]$	-0.171	-0.0856	0.0712	0.281	0.155	-0.159
	(0.187)	(0.200)	(0.223)	(0.174)	(0.185)	(0.211)
Mean pH	-0.120	-0.118	0.130	0.190	-0.149	0.0942
	(0.242)	(0.293)	(0.210)	(0.189)	(0.246)	(0.164)
N	285	285	285	291	291	291
Gender	Male	Male	Male	Female	Female	Female
Overall R^2	0.35	0.36	0.27	0.31	0.34	0.38

Notes: Robust standard errors clustered at district level in parentheses. Pollutants: FCOLI = faecal coliform, NIT = Nitrate-N + Nitrite-N, DO = Dissolved Oxygen. Explanatory variables not reported: Numerical variables such as Age, Height, Weight, "hours spent at school per week", "hours spend doing homework per week", "hours spent being tutored per week", "distance from school to home", "number of days the child spent disabled because of short-term morbidity in the last 30 days". Binary Variables such as Sex, "Rupees spent on books and uniform > Rs. 500", "household consumption per capita ≤ 25th percentile", "household consumption ≤ 50th percentile", "household consumption ≤ 75th percentile", "1 = water storage vessel available at home", "1 = water is purified at home though some mode of filtration or boiling", "1 = household members always wash hands after defaecation".

Explanation for table A6

We examine the impact of faecal coliform and Nitrate-N + Nitrite-N on boys versus girls in table A6. Most columns in this table do not present estimates that are statistically significant than zero. The comparison of the effect of $1[\overline{FCOLI} > limit]$ between column 9 and column 12 reveals a less than 0.095 standard deviation greater estimated effect on boys in writing tests, indicating that the effect of poor water quality is not necessarily uniform across genders. The analytical sample means in table A7 show that boys do not consistently outperform girls in tests across districts with unsafe pollutant levels. However, in districts with safe levels of pollutants, boys generally score higher. This disparity may be attributed to the fact that areas with safe faecal coliform levels are often rural, where girls may face more discrimination in educational investment and greater barriers to education.

Furthermore, we investigate whether girls in the sample received disproportionately lower investment in their education and if poor water quality and the adoption of water pollution mitigation technology affected them differently compared to boys. In tables A8 and A9, we include interaction terms to explore such possibilities. The estimated effects of interaction terms between being female and having faecal coliform levels above the safety level, the cost of books and uniforms incurred by the household, and whether the household has an indoor pump are examined in table A8. Although the interaction effects of being female and the cost of books and uniforms in 1000 Rupees are statistically significant in columns 7, 10 and 11, they cannot be fully interpreted due to the non-significant main effect of being female in these columns. We also explore interactions between being female and variables related to water availability and water purification in table A9. However, most estimated coefficients of these interaction terms are not statistically significant. It is possible that girls receive disproportionately less attention and investment from households regarding their education and are more affected by changes in water availability and exposure to water pollution, but our analyses in tables A8 and A9 do not find sufficient statistical evidence to confirm this.

Table A7. Comparison of mean male and female test scores by district-level prevalence of water pollution

		$1[\overline{FCOLI} > limit]$	$1[\overline{NIT} > limit]$	$0[\overline{FCOLI} \leq limit]$	$0[\overline{NIT} \leq limit]$
Sex					
Male	Reading	0.155	0.373	0.09	0.051
		(0.929)	(0.816)	(0.891)	(0.938)
	Maths	0.234	0.432	0.256	0.171
		(1.02)	(0.984)	(0.945)	(0.999)
	Writing	0.147	0.281	0.133	0.092
		(0.999)	(0.945)	(0.997)	(0.945)
Sex					
Female	Reading	0.177	0.346	0.025	0.59
		(0.929)	(0.828)	(0.948)	(0.962)
	Maths	0.194	0.265	0.042	0.112
		(1.007)	(0.943)	(0.964)	(1.014)
	Writing	0.212	0.302	0.103	0.139
		(0.993)	(0.910)	(1.01)	(1.026)

Notes: Reading, Writing, and Maths Test Scores have been Z-scored for the entire sample. Standard Deviations in parentheses. FCOLI = Faecal Coliform, NIT = Nitrate-N + Nitrite-N. Columns 1 and 2 show test score means for districts with unsafe levels of faecal coliform and Nitrate- N + Nitrite-N, respectively, while columns 3 and 4 show test score means for districts within safe levels of these pollutants.

Table A8. Differential effect of water pollution on girls

	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
Female	-0.00831	-0.124	0.0511	0.0716	-0.0269	-0.0163
	(0.0684)	(0.0987)	(0.0810)	(0.0844)	(0.0640)	(0.0758)
(1) $1[\overline{FCOLI} > limit]$	-0.231	-0.222	-0.197	-0.107	-0.0804	-0.547
	(0.0907)	(0.139)	(0.177)	(0.194)	(0.201)	(0.239)
Female × (1)	-0.0223	0.0465	-0.0283	-0.0919	-0.0658	0.0457
	(0.0806)	(0.128)	(0.109)	(0.112)	(0.101)	(0.103)
$1[\overline{NIT} > limit]$	0.0498	-0.0353	-0.133	0.0948	0.00750	-0.109
	(0.0744)	(0.0751)	(0.0872)	(0.260)	(0.131)	(0.228)
	(7) Reading Score	(8) Maths Score	(9) Writing Score	(10) Reading Score	(11) Maths Score	(12) Writing Score
Female	0.0274	-0.0458	0.0100	0.0631	0.0112	0.0343
	(0.0517)	(0.0497)	(0.0497)	(0.0928)	(0.102)	(0.0747)
(2) Cost of Books, Uniform	0.103	0.0665	0.0163	0.0940	0.0799	-0.0134
in 1000 Rupees.	(0.0291)	(0.0377)	(0.0363)	(0.0494)	(0.0543)	(0.0360)
Female × (2)	-0.0630	-0.0555	0.0257	-0.0806	-0.109	-0.0138
	(0.0293)	(0.0355)	(0.0330)	(0.0466)	(0.0490)	(0.0356)
$1[\overline{FCOLI} > limit]$	-0.243	-0.199	-0.211	-0.169	-0.133	-0.526
	(0.0833)	(0.101)	(0.155)	(0.169)	(0.168)	(0.223)
$1[\overline{NIT} > limit]$	0.0415	-0.0426	-0.130	0.0699	-0.0295	-0.117
	(0.0745)	(0.0756)	(0.0870)	(0.267)	(0.144)	(0.231)
	(13) Reading Score	(14) Maths Score	(15) Writing Score	(16) Reading Score	(17) Maths Score	(18) Writing Score
Female	-0.0317	-0.0907	0.0247	-0.0193	-0.0838	-0.00166
	(0.0401)	(0.0479)	(0.0442)	(0.0650)	(0.0836)	(0.0626)
(3) HH indoor piped water	0.0939	0.0292	0.134	0.252	0.0666	0.0534
	(0.115)	(0.0872)	(0.136)	(0.125)	(0.129)	(0.149)
Female × (3)	0.0593	-0.00682	0.0434	0.0222	-0.00904	0.143
	(0.0930)	(0.103)	(0.165)	(0.119)	(0.169)	(0.200)
$1[\overline{FCOLI} > limit]$	-0.244	-0.199	-0.213	-0.149	-0.113	-0.526
	(0.0821)	(0.101)	(0.155)	(0.154)	(0.163)	(0.214)
$1[\overline{NIT} > limit]$	0.0549	-0.0348	-0.128	0.160	0.0238	-0.0689
	(0.0730)	(0.0755)	(0.0850)	(0.232)	(0.122)	(0.214)
Observations	1147	1147	1147	576	576	576
Sample	All	All	All	River	River	River

Notes: Robust standard errors clustered at district level in parentheses. FCOLI = Faecal coliform and NIT = Nitrate-N + Nitrite-N.

Explanation of table A8

The regression specifications corresponding to the results in columns 1 to 18 include all explanatory variables used for the results in tables 2, 3 and 4 with additional controls for (1) whether respondent child receives scholarship for education, (2) year-round water availability (1=adequate, 0=inadequate), (3) drinking water storage vessel (1=the household has storage vessel, 0=none), (4) HH boils water to purify water (1=does, 0=does not), (5) frequency of washing hands after defaecation, (6) completed years of education, (7) Binary: Whether household has mobile phones, (8) Binary: whether household has computer, (9) short-term morbidity controls - the number of days the child was disabled in the last thirty days, the number of days the child showed certain symptoms like fever and coughing, and the amount of medical cost due to the short-term morbidity in the last thirty days, (10) school fees, whether the child gets free uniform (binary), costs of books, (11) Whether the household head considers the child's class teacher to be fair, parents' PTA participation (binary), child's teacher's gender, whether the child's admission to school was hard, frequency of child's teacher being absent at school. Lastly, district-level average short-term morbidity, state fixed effects, and survey month controls are also included to estimate each result.

Table A9. Differential effect of water pollution on girls (continued...)

	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
Female	0.0114	-0.141	-0.181	0.331	-0.493	-0.225
	(0.183)	(0.303)	(0.361)	(0.337)	(0.346)	(0.278)
(1) Water availability normal	-0.165	-0.0181	-0.0279	0.0272	0.124	-0.0852
the whole year = 1	(0.157)	(0.149)	(0.187)	(0.276)	(0.292)	(0.285)
Female × (1)	-0.0366	0.0511	0.219	-0.343	0.417	0.252
	(0.190)	(0.308)	(0.360)	(0.344)	(0.333)	(0.257)
$1[\overline{FCOLI} > limit]$	-0.242	-0.199	-0.212	-0.155	-0.115	-0.524
	(0.0834)	(0.100)	(0.157)	(0.164)	(0.164)	(0.224)
$1[\overline{NIT} > limit]$	0.0496	-0.0350	-0.132	0.0968	0.0147	-0.109
	(0.0746)	(0.0749)	(0.0875)	(0.262)	(0.130)	(0.227)
	(7) Reading Score	(8) Maths Score	(9) Writing Score	(10) Reading Score	(11) Maths Score	(12) Writing Score
Female	-0.107	-0.189	0.0986	0.00506	0.240	0.368
	(0.0705)	(0.164)	(0.209)	(0.211)	(0.425)	(0.265)
(2) Water availability normal	0.0909	0.106	-0.0727	-0.0132	-0.331	-0.362
during summer	(0.0873)	(0.158)	(0.211)	(0.236)	(0.417)	(0.304)
Female × (2)	-0.0630	-0.0555	0.0257	-0.0806	-0.109	-0.0138
	(0.0293)	(0.0355)	(0.0330)	(0.0466)	(0.0490)	(0.0356)
$1[\overline{FCOLI} > limit]$	-0.228	-0.187	-0.192	-0.118	-0.0948	-0.515
	(0.0840)	(0.103)	(0.157)	(0.159)	(0.163)	(0.215)
$1[\overline{NIT} > limit]$	0.0487	-0.0356	-0.137	0.0725	0.00707	-0.114
	(0.0747)	(0.0757)	(0.0849)	(0.259)	(0.121)	(0.226)
	(13) Reading Score	(14) Maths Score	(15) Writing Score	(16) Reading Score	(17) Maths Score	(18) Writing Score
Female	-0.107	-0.189	0.0986	0.00506	0.240	0.368
	(0.0705)	(0.164)	(0.209)	(0.211)	(0.425)	(0.265)
(3) HH purifies water	-0.0478	-0.0304	0.274	0.000347	0.461	0.296
	(0.111)	(0.157)	(0.194)	(0.217)	(0.302)	(0.346)
Female × (3)	0.0593	-0.00682	0.0434	0.0222	-0.00904	0.143
	(0.0930)	(0.103)	(0.165)	(0.119)	(0.169)	(0.200)
$1[\overline{FCOLI} > limit]$	-0.228	-0.187	-0.192	-0.118	-0.0948	-0.515
	(0.0840)	(0.103)	(0.157)	(0.159)	(0.163)	(0.215)
$1[\overline{NIT} > limit]$	0.0487	-0.0356	-0.137	0.0725	0.00707	-0.114
	(0.0747)	(0.0757)	(0.0849)	(0.259)	(0.121)	(0.226)
Observations	1143	1143	1143	573	573	573
Sample	All	All	All	River	River	River

Note: Robust standard errors clustered at district level in parentheses.

Explanation of table A9

The regression specifications corresponding to the results in columns 1 to 18 include all explanatory variables used for results in tables 2, 3 and 4 with additional controls for (1) whether respondent child receives scholarship for education, (2) year-round water availability (1=adequate, 0=inadequate), (3) drinking water storage vessel (1=the household has storage vessel, 0=none), (4) Whether the HH boils water to purify water (1=does, 0=does not), (5) frequency of washing hands after defaecation, (6) completed years of education, (7) Binary: Whether household has mobile phones, (8) Binary: Whether household has computer, (9) Short-term morbidity controls - the number of days the child was disabled in the last thirty days, the number of days the child showed certain symptoms like fever and coughing, and the amount of medical cost due to the short-term morbidity in the last thirty days, (10) school fees, whether the child gets free uniform (binary), costs of books, (11) Whether the household head considers the child's class teacher to be fair, parents' PTA participation (binary), child's teacher's gender, whether the child's admission to school was difficulty, frequency of child's teacher being absent at school. Lastly, district-level average short-term morbidity, state fixed effects, and survey month controls are also included to estimate each result.

Table A10. Differences in river pollution effect across states

Test	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score	(7) Reading Score	(8) Maths Score	(9) Writing Score
$1[\overline{FCOLI} > limit]$	-0.235	-0.203	-0.287	-0.0905	-0.0519	-0.448	0.192	0.180	-0.293
	(0.130)	(0.0820)	(0.184)	(0.175)	(0.164)	(0.0880)	(0.133)	(0.283)	(0.163)
$1[\overline{NIT} > limit]$	-0.0767	-0.131	0.219	–	–	–	-0.0343	0.0556	-0.0547
	(0.173)	(0.119)	(0.164)	–	–	–	(0.0734)	(0.166)	(0.0810)
$1[\overline{D.O.} < threshold]$	0.0336	-0.0674	-0.249	-0.485	-0.487	0.129	0.105	0.0888	-0.0112
	(0.186)	(0.153)	(0.197)	(0.118)	(0.127)	(0.0722)	(0.0998)	(0.167)	(0.106)
Mean pH	0.0122	-0.399	-0.0383	1.472	1.023	0.707	-0.125	-0.573	0.355
	(0.170)	(0.135)	(0.230)	(1.076)	(1.014)	(0.751)	(0.268)	(0.329)	(0.248)
N	422	422	422	347	347	347	367	367	367
States	UP	UP	UP	WB	WB	WB	BJ	BJ	BJ

Notes: Robust standard errors clustered at district level in parentheses. FCOLI = Faecal coliform and NIT = Nitrate-N + Nitrite-N. UP = Uttar Pradesh, BJ = Bihar and Jharkhand, and WB = West Bengal. The results in columns 1 to 9 are estimated by including the following control variables in the regression equation : Numerical variables such as Age, Sex, Height, Weight, "hours spent at school per week", "hours spend doing homework per week", "hours spent being tutored per week", "distance from school to home", "number of days the child spent disabled because of short-term morbidity in the last 30 days". Binary Variables such as Sex, "Rupees spent on books and uniform > Rs. 500", "household consumption per capita ≤ 25^{th} percentile", "household consumption ≤ 50th percentile", "household consumption ≤ 75th percentile", "1 = water storage vessel available at home", "1 = water is purified at home though some mode of filtration or boiling", "1 = household members always wash hands after defaecation".

Table A11. Robustness check with additional education-related and short-term morbidity-related controls

	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
$\overline{1[FCOLI > limit]}$	-0.0435	-0.196	-0.113	-0.297	-0.0372	-0.356
	(0.0671)	(0.141)	(0.0839)	(0.130)	(0.121)	(0.182)
$1[NIT > limit]$	-0.164	-0.0538	-0.00910	0.114	0.0259	0.0590
	(0.0792)	(0.171)	(0.0754)	(0.0952)	(0.0916)	(0.177)
$1[D.O. < threshold]$	-0.0319	0.0713	-0.0661	-0.00994	-0.156	-0.0722
	(0.0971)	(0.162)	(0.101)	(0.136)	(0.115)	(0.205)
Mean pH	0.0276	0.217	-0.237	-0.0533	-0.0313	0.175
	(0.0988)	(0.164)	(0.129)	(0.202)	(0.0932)	(0.191)
School distance	-0.0162	-0.0218	0.0199	0.0105	-0.00142	0.0180
	(0.0121)	(0.0176)	(0.0101)	(0.0154)	(0.0132)	(0.0191)
School hrs/week	-0.00590	-0.0138	-0.00482	-0.00288	-0.00991	-0.0192
	(0.00383)	(0.00447)	(0.00385)	(0.00451)	(0.00484)	(0.00567)
Homework hrs/week	0.0259	0.0277	0.0272	0.0276	0.0290	0.0311
	(0.00506)	(0.00750)	(0.00577)	(0.00684)	(0.00443)	(0.00616)
Pvt. Tuition hrs/week	0.0108	0.0187	0.0214	0.0249	0.0251	0.0329
	(0.00590)	(0.00893)	(0.00877)	(0.0112)	(0.00751)	(0.0119)
Cost Books, uniform	0.323	0.269	0.222	0.176	0.326	0.245
	(0.0570)	(0.0796)	(0.0539)	(0.0949)	(0.0554)	(0.0889)
STM days disabled†	0.00610	-0.0155	-0.00430	0.0126	-0.0272	-0.0258
	(0.0131)	(0.0163)	(0.0144)	(0.0246)	(0.00937)	(0.0210)
Indoor Piped Water	0.109	-0.0764	0.0510	-0.202	0.0267	-0.0852
	(0.0831)	(0.0960)	(0.0710)	(0.108)	(0.0842)	(0.0884)
1 to 4 years of Edu.	0.646	0.737	0.577	0.662	0.486	0.571
	(0.117)	(0.194)	(0.0888)	(0.0898)	(0.121)	(0.157)
N	1147	576	1147	576	1147	576
Sample	All	River	All	River	All	River
Overall R^2	0.30	0.27	0.27	0.19	0.26	0.29

Notes: Robust standard errors clustered at district level in parentheses. Pollutants: FCOLI = faecal coliform, NIT = Nitrate-N + Nitrite-N, DO = Dissolved Oxygen. †STM = short-term morbidity. Explanatory variables not reported :Numerical variables such as Age, Sex, Height, Weight, "household consumption per capita ≤ 25th percentile", "household consumption ≤ 50th percentile", "household consumption ≤ 75th percentile", "1 = water storage vessel available at home", "1 = water is purified at home though some mode of filtration or boiling", "1 = household members always wash hands after defaecation".

Explanation for table A11

Additional control variables used in estimating the results in table A11 are mentioned below: Private Tuition child receives hours/week, Short-term morbidity expenditure, 1 = Water availability is normal/adequate, Completed Years of schooling, 1 = Primary respondent of household owns mobile, 1 = Primary respondent of household uses a computer, Short-term morbidity - total cost, short-term morbidity - number of days ill in the last 30 days, Number of days with fever in the last 30 days, Number of days with cough in the last 30 days, Cost of treatment - travelling to health centre, cost of treatment - tests, medicines, miscellaneous. Descriptive Statistics of Teacher Quality variables are available in Description of these variables can be found in table A3.

Compared to the effects estimated in tables 2 and 4, the impact of $1[\overline{FCOLI} > limit]$ on reading and writing test scores is approximately 0.005 standard deviations lower in table A11. Moreover, column 1 in table A11 shows that the estimated effect of $1[\overline{NIT} > limit]$ is statistically significant for the full sample, whereas the effect of $1[\overline{FCOLI} > limit]$ on maths and writing test scores is not statistically significant.

Table A12. Robustness check with teaching quality controls

	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
$1[\overline{FCOLI} > limit]$	-0.0459	-0.150	-0.132	-0.294	-0.0382	-0.359
	(0.0864)	(0.174)	(0.0820)	(0.132)	(0.146)	(0.210)
$1[\overline{NIT} > limit]$	-0.161	-0.0611	-0.00779	0.0772	0.0259	0.0393
	(0.0817)	(0.157)	(0.0803)	(0.0664)	(0.102)	(0.156)
$1[\overline{D.O.} < threshold]$	0.0130	0.145	-0.0150	0.131	-0.115	0.0826
	(0.0881)	(0.163)	(0.0938)	(0.108)	(0.121)	(0.197)
Mean pH	0.0819	0.276	-0.227	-0.0393	-0.0350	0.224
	(0.0945)	(0.153)	(0.107)	(0.118)	(0.122)	(0.201)
T Fair	-0.146	-0.153	-0.163	-0.203	0.00382	0.0135
	(0.0863)	(0.112)	(0.0963)	(0.131)	(0.0886)	(0.156)
PTA Participation	0.0713	0.0471	0.138	0.277	0.0936	0.233
	(0.0773)	(0.0812)	(0.0687)	(0.0743)	(0.0802)	(0.0881)
T Biased	-0.0996	-0.184	0.0839	-0.0448	0.00904	0.438
	(0.117)	(0.213)	(0.1000)	(0.147)	(0.127)	(0.182)
Local T	0.152	0.212	0.175	0.211	0.174	0.167
	(0.0489)	(0.0855)	(0.0689)	(0.0838)	(0.0837)	(0.115)
Female T	0.0490	-0.0568	0.0438	0.0328	0.0421	-0.134
	(0.0434)	(0.0513)	(0.0634)	(0.0838)	(0.0738)	(0.0711)
T Attendance Regular	0.0824	0.170	0.0201	0.116	0.0494	0.101
	(0.0583)	(0.0917)	(0.0751)	(0.0826)	(0.0791)	(0.0784)
S Admission difficult	-0.0590	-0.134	0.101	0.0412	0.0649	-0.0115
	(0.0908)	(0.112)	(0.0966)	(0.161)	(0.0892)	(0.128)
N	1147	576	1147	576	1147	576
Sample	All	River	All	River	All	River

Notes: Robust standard errors clustered at district level in parentheses. T = teacher. S = School. PTA = Parent-Teacher Association. FCOLI = faecal coliform, NIT = Nitrate-N + Nitrite-N, DO = Dissolved Oxygen. The regression specifications corresponding the results in columns 1 to 6 include all explanatory variables used for results in table A11.

Table A13. Mixed-effects specifications - random intercepts

	(1) Reading Score	(2) Reading Score	(3) Reading Score	(4) Reading Score	(5) Reading Score	(6) Reading Score
$1[\overline{FCOLI} > limit]$	-0.109	-0.582	-0.234	-0.0290	-0.288	-0.0170
	(0.0777)	(0.308)	(0.112)	(0.135)	(0.161)	(0.114)
$1[\overline{NIT} > limit]$	-0.129	-0.0563	-0.0650	-0.104	-0.00507	-0.0621
	(0.0805)	(0.196)	(0.0974)	(0.115)	(0.146)	(0.124)
	(7) Maths Score	(8) Maths Score	(9) Maths Score	(10) Maths Score	(11) Maths Score	(12) Maths Score
$1[\overline{FCOLI} > limit]$	-0.145	-0.669	-0.322	0.106	-0.192	-0.0420
	(0.0874)	(0.267)	(0.119)	(0.197)	(0.180)	(0.121)
$1[\overline{NIT} > limit]$	-0.0444	0.112	0.0868	-0.000282	0.163	-0.166
	(0.0925)	(0.178)	(0.104)	(0.177)	(0.164)	(0.131)
	(13) Writing Score	(14) Writing Score	(15) Writing Score	(16) Writing Score	(17) Writing Score	(18) Writing Score
$1[\overline{FCOLI} > limit]$	-0.157	-0.341	-0.355	-0.00710	-0.378	-0.103
	(0.0944)	(0.278)	(0.150)	(0.111)	(0.175)	(0.120)
$1[\overline{NIT} > limit]$	0.0878	-0.0870	0.119	0.0242	0.112	0.0907
	(0.102)	(0.186)	(0.134)	(0.0945)	(0.160)	(0.128)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.

Notes: Robust standard errors clustered at district level in parentheses. FCOLI = Faecal coliform, NIT = Nitrate-N + Nitrite-N, GW = groundwater. The regression specifications corresponding the results in columns 1 to 18 include all explanatory variables used for results in tables 2, 3 and 4.

Table A14. Baseline model estimated with random intercepts at two levels, district and village

Test	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
$1[\overline{FCOLI} > limit]$	-0.0412	-0.161	-0.142	-0.281	-0.0839	-0.365
	(0.0788)	(0.148)	(0.0865)	(0.133)	(0.0989)	(0.173)
$1[\overline{NIT} > limit]$	-0.134	-0.0196	-0.00644	0.110	0.102	0.0785
	(0.0777)	(0.135)	(0.0862)	(0.118)	(0.106)	(0.158)
$1[\overline{D.O.} < threshold]$	-0.0448	0.0353	-0.0536	0.0204	-0.144	-0.0254
	(0.0754)	(0.145)	(0.0823)	(0.130)	(0.0935)	(0.173)
Mean pH	0.0618	0.247	-0.198	-0.0956	-0.0712	0.0944
	(0.104)	(0.192)	(0.114)	(0.161)	(0.131)	(0.223)
N	1147	576	1147	576	1147	576
Districts	All	River	All	River	All	River

Notes: Robust standard errors clustered at district level in parentheses. "Uttarakhand" is the reference State to the State dummy variables. The results in columns 1 to 6 are estimated by including all the control variables used for table A16 results except Mean District-level Short-term Morbidity.

Explanation for table A14

Table A14 presents findings from a two-level random-intercept model used for additional robustness checks. This model accommodates the hierarchical structure of the data by recognising villages and/or neighbourhoods as the primary levels within which the surveyed children are 'nested.' By adopting this approach, we can estimate the effects of $1[\overline{FCOLI} > limit]$ and $1[\overline{NIT} > limit]$ on the dependent variable, considering potential intra-group correlations within each village or neighbourhood. The results shown in table A14 indicate that the estimated effects of exceeding safe levels of faecal coliform and nitrate-nitrite are consistent and statistically significant in the two-level random-intercept model.

Table A15. Baseline regression - random intercepts at three levels (district, village, and household)

Test	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
$1[\overline{FCOLI} > limit]$	-0.0452	-0.163	-0.131	-0.275	-0.0823	-0.361
	(0.0800)	(0.147)	(0.0887)	(0.135)	(0.100)	(0.170)
$1[\overline{NIT} > limit]$	-0.134	-0.0187	-0.0167	0.108	0.0904	0.0702
	(0.0787)	(0.134)	(0.0885)	(0.120)	(0.107)	(0.156)
$1[\overline{D.O.} < threshold]$	-0.0314	0.0389	-0.0548	0.0210	-0.131	-0.00979
	(0.0767)	(0.144)	(0.0846)	(0.132)	(0.0951)	(0.170)
Mean pH	0.0520	0.241	-0.187	-0.0915	-0.0771	0.103
	(0.105)	(0.190)	(0.117)	(0.164)	(0.132)	(0.218)
N	1147	576	1147	576	1147	576
Districts	All	River	All	River	All	River

Notes: Robust standard errors clustered at district level in parentheses. DO = Dissolved Oxygen. "Uttarakhand" is the reference State to the State dummy variables. The results in columns 1 to 6 are estimated by including all the control variables used for table A16 results except Mean District-level short-term morbidity.

Explanation for table A15

Table A15 extends the multilevel analysis presented in table A14 by adding households as an additional level, creating a three-level random-intercept model. This more granular approach allows for the consideration of variation both within and between households, villages, and neighbourhoods, providing a comprehensive view of the data's nested structure. The results in table A15 offer insight into the impact of water pollutants, with specific attention to faecal coliform and nitrate-nitrite levels that surpass safe thresholds. The estimated effects remain statistically robust even when the household level is included.

Table A16. Robustness check with district-level morbidity as control

Test	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
$1[\overline{FCOLI} > limit]$	-0.0120	-0.147	-0.106	-0.295	0.0413	-0.303
	(0.0832)	(0.172)	(0.0813)	(0.128)	(0.130)	(0.191)
$1[\overline{NIT} > limit]$	-0.154	-0.0554	0.00218	0.0854	0.0347	0.0443
	(0.0774)	(0.159)	(0.0824)	(0.0774)	(0.0896)	(0.159)
$1[\overline{D.O.} < threshold]$	-0.0548	0.0341	-0.0750	0.0295	-0.216	-0.0759
	(0.0948)	(0.159)	(0.0920)	(0.116)	(0.117)	(0.209)
Mean pH	0.0422	0.195	-0.249	-0.0887	-0.0561	0.202
	(0.0882)	(0.141)	(0.114)	(0.145)	(0.0985)	(0.167)
Mean Morbidity	-0.00502	0.00481	-0.00852	-0.00318	-0.0710	-0.0874
	(0.0155)	(0.0298)	(0.0181)	(0.0305)	(0.0175)	(0.0323)
N	1147	576	1147	576	1147	576
Districts	All	River	All	River	All	River

Notes: Robust standard errors clustered at district level in parentheses. DO = Dissolved Oxygen T = teacher. S = School. PTA = Parent-Teacher Association. The results in columns 1 to 6 are estimated by including all the control variables used for table A12 results.

Explanation of table A16

The channel through which water pollution affects educational outcomes is the deterioration of the child's health, primarily followed by a decrease in cognitive abilities. We only control for short-term morbidity and do not have any measure of long-term morbidity. As a control for the district-wide morbidity level, we create a measure of the district-average short-term morbidity of the children within a district. As a control variable, it should account for the average effect of district-level short-term morbidity caused by water pollution. Controlling for the channel of short-term morbidity may leave the secondary channel of a decrease in cognitive abilities through long-term morbidity or recurring short-term morbidity open. We observe that short-term morbidity, when included in the model, weakens the effect of 1[FCOLI > limit] on writing test scores in districts where water was monitored (column 6 in table A16, compared to column 6 in table A11). However, the effect of 1[FCOLI > limit] on reading scores in districts along rivers and that of 1[NIT > limit] on reading scores in the all-district sample remain statistically robust.

Table A17. Robustness check with state fixed effects

Test	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
$1[\overline{FCOLI} > limit]$	-0.212	-0.202	-0.171	-0.186	-0.254	-0.545
	(0.0911)	(0.189)	(0.0966)	(0.161)	(0.147)	(0.245)
$1[\overline{NIT} > limit]$	-0.0774	-0.0146	0.0345	0.0973	0.126	0.152
	(0.0721)	(0.138)	(0.0909)	(0.0905)	(0.0853)	(0.108)
$1[\overline{D.O.} < threshold]$	-0.0366	-0.00498	-0.0703	0.00795	-0.132	-0.166
	(0.0799)	(0.138)	(0.0987)	(0.119)	(0.108)	(0.138)
Mean pH	0.0914	0.206	-0.223	-0.115	0.134	0.234
	(0.123)	(0.143)	(0.163)	(0.161)	(0.149)	(0.142)
Uttar Pradesh	0.358	0.215	0.739	0.633	0.484	0.307
	(0.0989)	(0.156)	(0.144)	(0.172)	(0.136)	(0.171)
Bihar	0.116	0.0396	0.809	0.862	0.0904	-0.0496
	(0.162)	(0.252)	(0.162)	(0.175)	(0.150)	(0.241)
West Bengal	0.555	0.168	1.109	0.814	0.719	0.666
	(0.179)	(0.232)	(0.172)	(0.194)	(0.162)	(0.182)
Jharkhand	0.00586	-	0.611	-	0.424	-
	(0.238)	-	(0.227)	-	(0.201)	-
N	1147	576	1147	576	1147	576
Districts	All	River	All	River	All	River

Notes: Robust standard errors clustered at district level in parentheses. DO = Dissolved Oxygen "Uttarakhand" is the reference State to the State dummy variables. The results in columns 1 to 6 are estimated by including all the control variables used for table results except Mean district-level short-term morbidity.

Explanation for table A17

If state-specific policies in pollution control and education provision are connected through the channel of quality governance, then the pollution treatment effects would be underestimated. So, we add state fixed effects to our model and estimate the results in table A17. Table A17 presents the change in the coefficient estimates of unsafe levels of faecal coliform and Nitrate-N + Nitrite-N after the inclusion of state-fixed effects. We can see that the effect of $1[\overline{NIT} > limit]$ in column 1 of table A17 - compared to column 1 in table A16 - is indistinguishable from zero. On the other hand, the estimated effect of $1[\overline{FCOLI} > limit]$ remains statistically significant for writing (columns 3 and 6), reading (column 1), and math (column 5). The state dummy for Jharkhand is not identified because the only district in Jharkhand in our analytical sample, Palamu, is not situated by a river. Overall, we find the math score to be less responsive to pollution. According to Ashraf (2020) and Babu (2012), and due to resource constraints faced by rural and/or public schools and poorer sections of the urban population, teaching math compared to other subjects is more difficult. This difficulty likely results in low variations in the math test score and therefore sees little effect of factors like river pollution.

Table A18. Checking for seasonality

State_ID×District Mean Morbidity×Survey Month

	(1) Reading Score	(2) Reading Score	(3) Reading Score	(4) Reading Score	(5) Reading Score	(6) Reading Score
$1[\overline{FCOLI} > limit]$	-0.303	-0.0974	0.0433	-0.302	0.00431	-0.722
	(0.0831)	(0.410)	(0.201)	(0.161)	(0.198)	(0.113)
$1[\overline{NIT} > limit]$	-0.0542	-0.346	0.404	0.0162	0.379	-0.172
	(0.0731)	(0.200)	(0.283)	(0.0756)	(0.317)	(0.0777)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.
	(7) Maths Score	(8) Maths Score	(9) Maths Score	(10) Maths Score	(11) Maths Score	(12) Maths Score
$1[\overline{FCOLI} > limit]$	-0.246	-0.132	-0.230	-0.156	-0.235	-0.319
	(0.106)	(0.274)	(0.116)	(0.156)	(0.129)	(0.116)
$1[\overline{NIT} > limit]$	0.0115	-0.181	0.245	0.0943	0.185	-0.0665
	(0.0788)	(0.134)	(0.532)	(0.0761)	(0.138)	(0.115)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.
	(13) Writing Score	(14) Writing Score	(15) Writing Score	(16) Writing Score	(17) Writing Score	(18) Writing Score
$1[\overline{FCOLI} > limit]$	-0.160	-0.181	-0.437	-0.0540	-0.436	0.0596
	(0.143)	(0.349)	(0.205)	(0.115)	(0.203)	(0.162)
$1[\overline{NIT} > limit]$	0.0698	0.101	0.0529	0.0531	-0.00564	0.164
	(0.110)	(0.186)	(0.258)	(0.0845)	(0.254)	(0.137)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.

Notes: Robust standard errors clustered at district level in parentheses. The regression specifications corresponding the results in columns 1 to 18 include all explanatory variables used for results in tables 2, 3, and 4 with additional controls for (1) whether respondent child receives scholarship for education, (2) year-round water availability (1=adequate, 0=inadequate), (3) drinking water storage vessel (1=the household has storage vessel, 0=none), (4) Whether the HH boils water to purify water (1=does, 0=does not), (5) frequency of washing hands after defaecation, (6) completed years of education, (7) Binary: Whether household has mobile phones, (8) Binary: Whether household has computer, (9) Short-term morbidity controls - the number of days the child was disabled in the last thirty days, the number of days the child showed certain symptoms like fever and coughing, and the amount of medical cost due to the short-term morbidity in the last thirty days, (10) school fees, whether the child gets free uniform (binary), costs of books, (11) Whether the household head considers the child's class teacher to be fair, parents' PTA participation (binary), child's teacher's gender, whether the child's admission to school was difficulty, frequency of child's teacher being absent at school.

Explanation for table A18

As the data used in this paper is cross-sectional, there is little opportunity to detect and control for seasonality for multiple years. table A18 shows how the effect of $1[\overline{FCOLI} > limit]$ changes in the presence of extensive controls for seasonality. Therefore, we include interaction terms - State ID × District Mean Morbidity × Survey Month - in the empirical model and present the results in table A18. These interaction terms control for survey-month specific variations in district-level average child short-term morbidity 'in the last thirty days from the date of the survey month' by states. Faecal coliform effects for reading and Maths in the first column (full sample) of table A18 are statistically significant after including these interaction terms. These interaction terms are viable controls under the assumption that within a state, district-level average short-term morbidity in different months remain unchanged over 2011 and 2012 - the two years when the last wave of the survey took place - and that in the months when the survey was not conducted - July, August, and September - seasonal variation in district-level average short-term morbidity does not exceed that of the previous and subsequent months.

Table A19. Robustness checks: proxying for air pollution using PM2.5 in 2012

	(1) Reading Score	(2) Reading Score	(3) Reading Score	(4) Reading Score	(5) Reading Score	(6) Reading Score
$1[\overline{FCOLI} > limit]$	-0.245	-1.088	-0.0874	-0.265	-0.160	-0.448
	(0.0868)	(0.805)	(0.157)	(0.164)	(0.187)	(0.129)
$1[\overline{NIT} > limit]$	0.0579	0.307	0.0410	-0.0132	0.124	0.0218
	(0.0833)	(0.262)	(0.271)	(0.0775)	(0.265)	(0.0721)
PM2.5 in 2012	-0.00141	-0.0144	0.0119	0.00686	-0.00764	0.00748
	(0.00371)	(0.0193)	(0.01610)	(0.00768)	(0.00502)	(0.00830)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.
	(7) Maths Score	(8) Maths Score	(9) Maths Score	(10) Maths Score	(11) Maths Score	(12) Maths Score
$1[\overline{FCOLI} > limit]$	-0.209	0.769	-0.0521	-0.236	-0.122	-0.320
	(0.105)	(1.125)	(0.160)	(0.168)	(0.179)	(0.148)
$1[\overline{NIT} > limit]$	0.00927	-0.0162	-0.0194	-0.0437	0.0474	0.000753
	(0.0828)	(0.275)	(0.117)	(0.117)	(0.135)	(0.0892)
PM2.5 in 2012	-0.00772	0.0297	0.00138	-0.0121	-0.0114	-0.00720
	(0.00326)	(0.0274)	(0.00673)	(0.00775)	(0.00471)	(0.00612)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.
	(13) Writing Score	(14) Writing Score	(15) Writing Score	(16) Writing Score	(17) Writing Score	(18) Writing Score
$1[\overline{FCOLI} > limit]$	-0.213	-0.937	-0.472	-0.00802	-0.525	-0.0339
	(0.158)	(1.316)	(0.229)	(0.118)	(0.228)	(0.185)
$1[\overline{NIT} > limit]$	-0.126	-0.279	-0.156	-0.0586	-0.102	-0.241
	(0.101)	(0.379)	(0.207)	(0.0705)	(0.228)	(0.112)
PM2.5 in 2012	-0.00134	0.00350	0.0130	0.000578	-0.00293	0.000297
	(0.00408)	(0.0307)	(0.01732)	(0.00556)	(0.00534)	(0.00801)
N	1147	206	532	769	576	738
Sample	All	Lake	Ganges	GW	River	Trib.

Note: Robust standard errors clustered at district level in parentheses.

Explanation for table A19

Table A19 provides results illustrating the effects of unsafe levels of faecal coliform and Nitrate-N + Nitrite-N, as well as PM2.5 in 2012. The regression specifications corresponding the results in columns 1 to 18 include all explanatory variables used for results in tables 2, 3 and 4 with additional controls for (1) whether respondent child receives scholarship for education, (2) year-round water availability (1=adequate, 0=inadequate), (3) drinking water storage vessel (1=the household has storage vessel, 0=none), (4) Whether the HH boils water to purify water (1=does, 0=does not), (5) frequency of washing hands after defaecation, (6) completed years of education, (7) Binary: Whether household has mobile phones, (8) Binary: Whether household has computer, (9) Short-term morbidity controls - the number of days the child was disabled in the last thirty days, the number of days the child showed certain symptoms like fever and coughing, and the amount of medical cost due to the short-term morbidity in the last thirty days, (10) school fees, whether the child gets free uniform (binary), costs of books, (11) Whether the household head considers the child's class teacher to be fair, parents' PTA participation (binary), child's teacher's gender, whether the child's admission to school was difficulty, frequency of child's teacher being absent at school. Lastly, district-level average short-term morbidity, state fixed effects, and survey month controls are also included to estimate each result.

After the inclusion of PM2.5 as a control variable, the impact of unsafe levels of faecal coliform on reading scores remains statistically significant for the full sample (column 1, table A19) and the sample of districts along tributaries (column 6, table A19). Moreover, the effect of unsafe levels of faecal coliform was also found to be statistically significant on writing scores (columns 15 and 17, table A19) and on maths scores (columns 7 and 12). PM2.5 itself exhibited a statistically significant negative effect only on maths test scores in columns 7 and 11.

Table A20. The effect of district-level mean faecal coliform instrumented by mean faecal coliform in the immediate upstream district

All	(1) Reading Score	(2) Maths Score	(3) Writing Score	(4) Reading Score	(5) Maths Score	(6) Writing Score
Model	RE	RE	RE	RE-IV	RE-IV	RE-IV
Mean FCOLI	-0.0991	-0.0722	0.0308	-0.321	0.0445	0.0855
	(0.0505)	(0.0377)	(0.0256)	(0.147)	(0.0976)	(0.0791)
Mean NIT	-0.384	-0.276	-0.368	-0.352	-0.293	-0.376
	(0.277)	(0.221)	(0.138)	(0.390)	(0.180)	(0.135)
N	871	871	871	871	871	871
River	(7) Reading Score	(8) Maths Score	(9) Writing Score	(10) Reading Score	(11) Maths Score	(12) Writing Score
Model	RE	RE	RE	RE-IV	RE-IV	RE-IV
Mean FCOLI	-0.00365	-0.000807	-0.000521	-0.000751	-0.0000480	0.000206
	(0.000735)	(0.000862)	(0.000904)	(0.000199)	(0.000224)	(0.000284)
Mean NIT	-1.681	-0.549	-0.483	-1.602	-0.604	-0.460
	(0.269)	(0.316)	(0.317)	(0.233)	(0.266)	(0.276)
N	460	460	460	460	460	460
Tributary	(13) Reading Score	(14) Maths Score	(15) Writing Score	(16) Reading Score	(17) Maths Score	(18) Writing Score
Model	RE	RE	RE	RE-IV	RE-IV	RE-IV
Mean FCOLI	-0.716	-0.819	-0.0919	-0.717	-0.819	-0.0919
	(0.362)	(0.446)	(0.402)	(0.362)	(0.446)	(0.402)
Mean NIT	0.249	0.668	-0.296	0.249	0.668	-0.296
	(0.444)	(0.429)	(0.441)	(0.444)	(0.429)	(0.441)
N	567	567	567	567	567	567

Notes: Robust standard errors clustered at district level in parentheses. MeanFCOL = district-average faecal coliform MPN/100 ml, in millions. MeanNIT = district-average Nitrate-N + Nitrite-N mg/l. The results in columns labelled "RE-IV" utilize the district-average faecal coliform amount, which is instrumented by the average level of faecal coliform from its upstream neighbouring district. MeanFCOL>2500 MPN per 100 ml is the safety limit for faecal coliform. Mean NITRATE- N+ NITRITE-N are measured in milligrams per litre of water.

Explanation for table A20

Table A20 shows the effect of district-level mean faecal coliform instrumented by mean faecal coliform in the immediate upstream district. The regression specifications corresponding the results in columns 1 to 18 include all explanatory variables used for results in tables 2, 3 and 4 with additional controls for (1) whether respondent child receives scholarship for education, (2) year-round water availability (1=adequate, 0=inadequate), (3) drinking water storage vessel (1=the household has storage vessel, 0=none), (4) Whether the HH boils water to purify water (1=does, 0=does not), (5) frequency of washing hands after defaecation, (6) child's years of schooling, (7) binary: Whether household has mobile phones, (8) binary: Whether household has computer, (9) Short-term morbidity controls - the number of days the child was disabled in the last thirty days, the number of days the child showed certain symptoms like fever and coughing, and the amount of medical cost due to the short-term morbidity in the last thirty days, (10) school fees, whether the child gets free uniform (binary), costs of books, (11) Whether the household head considers the child's class teacher to be fair, (12) whether parents participate in PTA meetings (binary), (13) child's teacher's gender, (14) binary: whether the child's admission to school was difficulty, (15) frequency of child's teacher being absent at school.

Impact of upstream pollution on downstream districts: an IV approach

Pollution in upstream areas can often have significant consequences for downstream environments and communities, but the factors influencing this pollution upstream are reasonably independent from the factors influencing negative consequences downstream. Later in Section 4, we discuss some results where downstream district pollution is instrumented with upstream district pollution. For this instrumentation scheme, we trace the course of the river and identify one neighbouring upstream district for each downstream district. We only select the upstream district that is also adjacent or sharing administrative borders. This process is straightforward because, in the Ganges basin, most rivers and tributaries flow from northwest to southeast. The sample used in this particular analysis excludes some districts whose upstream counterpart was not included in IHDS.

We introduce an additional model to tackle the concern regarding the extent to which the impact of water quality was influenced by pollution originating from upstream districts. In this model, we employ district-level pollution as an instrumental variable, using the pollution levels of districts upstream and those sharing borders as counterparts. A comparable strategy is adopted by Do *et al. (*2018) to untangle localized unobserved variables from measurements of river water pollution. It is worth noting that pollution originating in upstream regions can negatively impact downstream environments and communities. The factors influencing this upstream pollution are generally independent of the observed and unobserved factors that con- tribute to pollution downstream. Therefore, for this analysis, we update our initial model as follows

$$Z_{ik} = \alpha_{ik} + \widehat{MeanFCOLI}_k + \widetilde{\boldsymbol{W}}'\Theta + \boldsymbol{X}'\Gamma + \widetilde{\chi_k} + \epsilon_{ik} \quad (3)$$

$$MeanFCOLI_k = \beta_1 + \widetilde{\boldsymbol{W}}'\Psi + \boldsymbol{X}'\Delta + upstream\ MeanFCOLI_{ik} + \omega_k + u_{ik} \quad (4)$$

With the assumption that $E(upstream\ Mean\ FCOLI\ |\omega_k)$=0, $\widetilde{\boldsymbol{W}}$ is the vector of water quality indicators except faecal coliform. Testing the validity of instruments becomes more complex with binary endogenous variables. Standard over-identification tests and tests for endogeneity might not apply in a straightforward way. Therefore, we use the district-level mean faecal coliform measures, $MeanFCOLI_k$ and $MeanFCOLI_k$, instead of binary measures. Unlike the binary measures of unsafe levels of faecal coliform, estimated and predicted $\widehat{MeanFCOLI}_k$ cannot be interpreted in a way that makes economic sense. As an additional robustness check, this exercise serves to show if faecal coliform instrumented by its upstream measure still affects test scores when it is disentangled from the district-level unobserved factors through instrumentation. For this development, we do lose some observations because not every district has an adjacent counterpart district along the same river that has also been included in the IHDS survey by Desai and Vanneman (2012). For this part, we use random-effects generalised 2SLS methods to estimate the instrumented model.

For accuracy, we instrument the district's faecal coliform level only with that of a border-sharing upstream district. For this exercise, we retain fewer observations compared to our primary analysis. We did not find data for upstream and border-sharing districts of 2 districts in Bihar, 2 districts in West Bengal, 1 district in Uttarakhand, and 3 districts in Uttar Pradesh. This choice is made considering that pollution generated by more distant districts tends to decay to some extent, potentially weakening the anticipated effect on downstream pollution levels. This portion of the analysis only discusses results from the instrumentation of district- average level of faecal coliform and not district-average level of Nitrate-N + Nitrite-N due to little variation between upstream and downstream levels of Nitrate-N + Nitrite-N. Overall, we have seen little effect of Nitrate-N + Nitrite-N on test scores; therefore, an additional analysis with the upstream district-average level of Nitrate-N + Nitrite-N as an instrument has been skipped in this section.

The outcomes of this instrumental variable (IV) analysis are outlined in table A20. We present the results from both the random-effects model and the random-effects model with instrumentation side by side for easy comparison. The results in columns labelled RE-IV are estimated using generalised 2SLS random-effects methods. Results in columns 1 to 6 are based on the all-district sample, columns 7 to 12 for districts along rivers, and columns 12 to 18 for districts along tributaries. Applying upstream faecal coliform measure as an instrument reveals a statistically significant impact of the district-average faecal coliform level on reading scores in three samples: all districts, river-adjacent districts, and tributary-adjacent districts. Additionally, we observe that instrumented $MeanFCOLI$ has some weak effect on maths test scores in the all-district sample and the 'tributaries' sample. However, we do not observe any statistically significant effect of instrumented $MeanFCOLI$ on Writing Score, which might result from using smaller samples where the treatment $MeanFCOLI$ varies only at the district level.

Figure A1. Five states of India within the Ganges Basin.

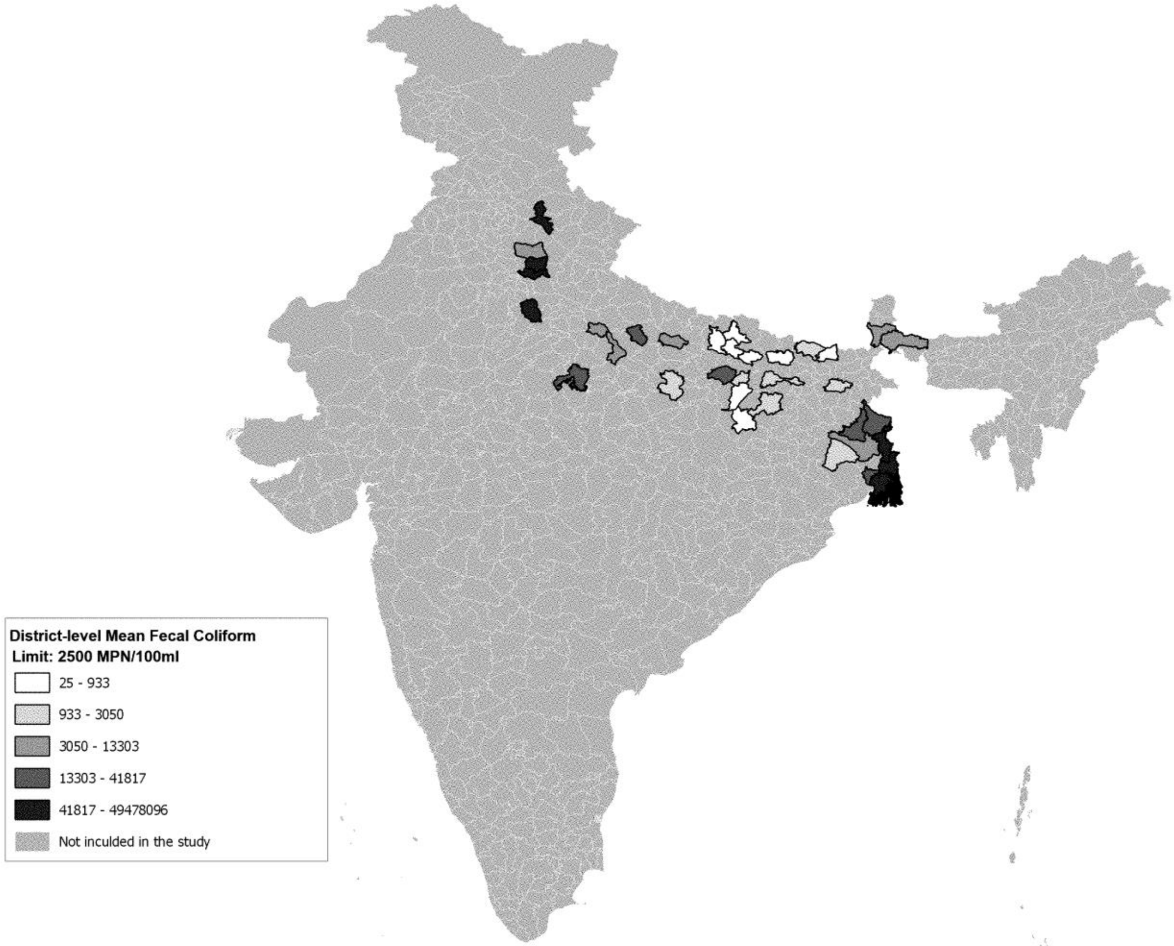

Figure A2. Mean faecal Coliform (MPN/100 ml).

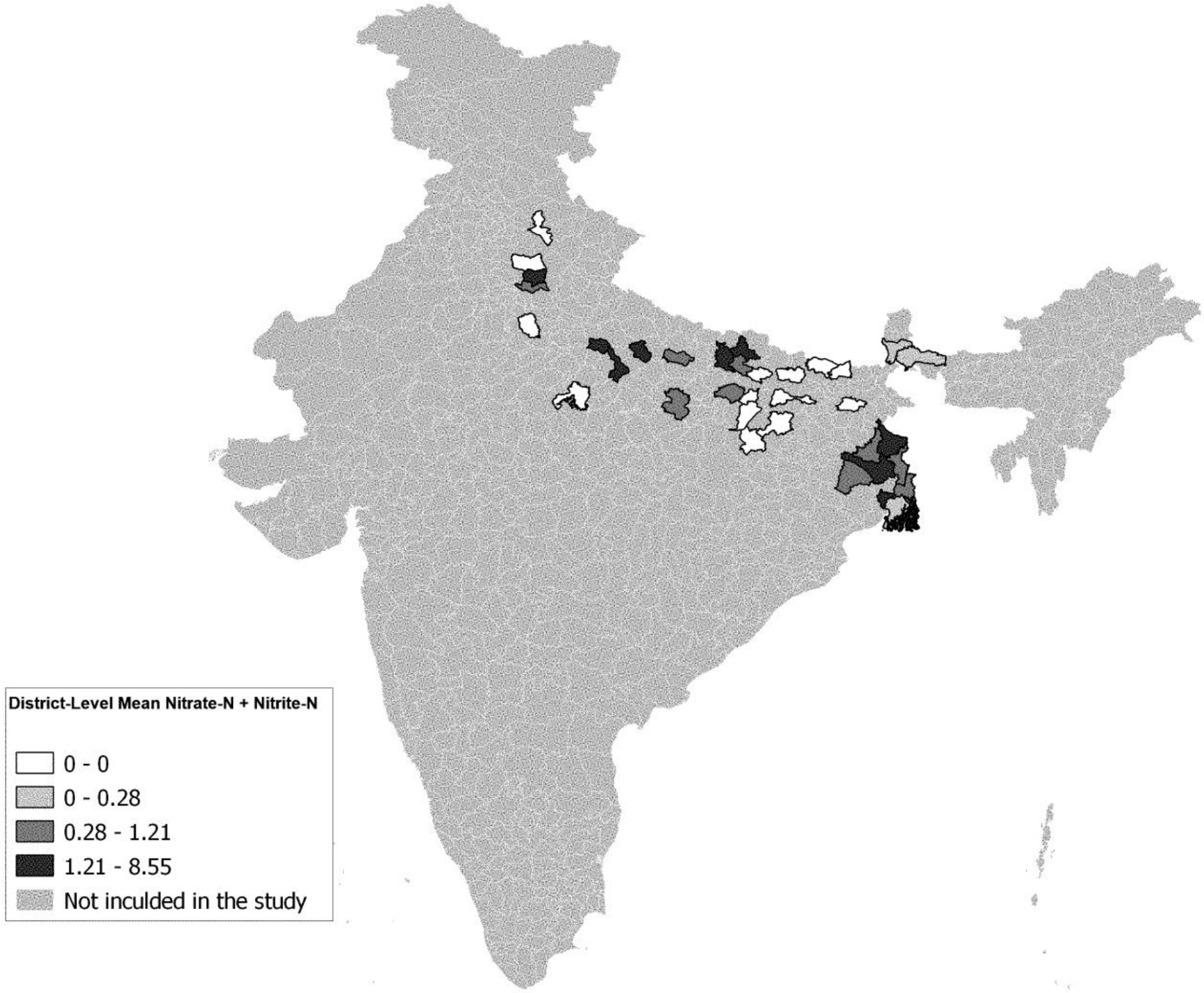

Figure A3. Mean Nitrate-N+Nitrite-N (mg/l).

References

Ashraf A (2020) Challenges and possibilities in teaching and learning of calculus: a case study of India. *Journal for the Education of Gifted Young Scientists* **8**, 407–433.

Babu S (2012) Indigenous traditions and the colonial encounter: a historical perspective on mathematics education in India. *Mathematics Education in India: Status and Outlook*, 37–62.

Desai S and Vanneman R (2012) India Human Development Survey-II (IHDS-II), 2011-12. Inter-university Consortium for Political and Social Research [distributor], 2018-08-08. Available at https://doi.org/10.3886/ICPSR36151.v6

Do QT, Joshi S and Stolper S (2018) Can environmental policy reduce infant mortality? Evidence from the Ganga pollution cases. *Journal of Development Economics* **133**, 306–325.

Pratham (2021) Pratham. Available at http://www.pratham.org