I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16863

Still Waters Run Deep: Groundwater Contamination and Education Outcomes in India

Khushboo Aggarwal
Rashmi Barua
Marian Vidal-Fernandez

MARCH 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16863

Still Waters Run Deep: Groundwater Contamination and Education Outcomes in India

Khushboo Aggarwal
Christ University

Rashmi Barua
CITD and JNU

Marian Vidal-Fernandez
University of Sydney and IZA

MARCH 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Still Waters Run Deep: Groundwater Contamination and Education Outcomes in India*

We investigate the impact of groundwater contamination on educational outcomes in India. Our study leverages variations in the geographical coverage and timing of construction of safe government piped water schemes to identify the effects of exposure to contaminants. Using self-collected survey data from public schools in Assam, one of the most groundwater-contaminated regions in India, we find that prolonged exposure to unsafe groundwater is associated with increased school absenteeism, grade retention, and decreased test scores and Cumulative Grade Point Average (CGPA). To complement our findings and to study the effect of one such contaminant, arsenic, we use a large nationally representative household survey. Using variations in soil textures across districts as an instrument for arsenic concentration levels we find that exposure to arsenic beyond safe threshold levels is negatively associated with school attendance.

JEL Classification: I15, I25, F63

Keywords: water contamination, education, India

Corresponding author:
Marian Vidal-Fernández
School of Economics
University of Sydney
NSW 2006
Sydney
Australia

E-mail: m.vidal-fdez@sydney.edu.au

* We are grateful to Anant Khanikar, Nipendra Sharma, Madhurjya Saikia, Ajoy Lahon, Runti Chowdhury, and Dr. Ratul Mahanta for support in the initial stages of this project. Thanks to Sultana Khan who was an excellent project coordinator and the entire team of research assistants in Jorhat. We thank participants in seminars at Jawaharlal Nehru University, Indian Statistical Institute, the 2018 ISI Conference on Economic Growth and Development, the CECFEE 2019 conference in Tezpur, Assam, the AASLE 2021, the 2022 AEDE conference at Porto and the European Winter Meeting of the Econometric Society (EWMES2023) in Manchester. Marian Vidal-Fernandez and Rashmi Barua gratefully acknowledges funding from the University of Sydney's India Development Fund. IERB JNU Ref. No. 2018/174.

1. Introduction

Across the world, more than 1,800 children under the age of five die every day from unsafe water consumption and inadequate sanitation (UNICEF, 2006). In India, where 67% of groundwater is already heavily contaminated with arsenic, fluoride, and nitrates, groundwater quality is becoming a pressing issue (Central Ground Water Board, 2022). While bacteria are more prevalent in surface water than groundwater, industrial and agricultural activities can worsen soil features that make heavy metals, such as naturally-occurring sulphates, iron, fluorides, nitrogen, chlorides, and arsenic, more abundant in groundwater. Moreover, climate change is reducing the rate at which rainwater seeps underground, gradually increasing the concentration of toxins in groundwater (McArthur et al., 2001).

A handful of papers have analysed the link between drinking contaminated groundwater on child health in developing countries.[1] Kile et al. (2016) find that mothers who drank arsenic contaminated water during pregnancy were more likely to give birth to low-weight infants. Brainerd and Menon (2014) exploit seasonal variation in exposure to fertilisers in groundwater during pregnancy and find a negative impact on infant and child health outcomes in India.[2]

While the effect of drinking contaminated water on a range of health outcomes, including growth outcomes among children, is well documented,[3] lesser is known about its causal effects on school absenteeism and academic achievement among children. This is an important question for public policy, because birth and health outcomes during early childhood play a crucial role in determining education and adult economic outcomes.[4] Moreover, school absenteeism has been shown to be a strong determinant of educational performance, achievement, and adult income (Cattan et al. 2023).

The consumption of contaminated groundwater impacts educational outcomes through its adverse effect on health of already vulnerable children. First, prolonged exposure to chemicals such as arsenic and lead has been linked to cognitive delays among children, particularly among those who suffer an additional

[1] A wide body of literature has studied the adverse impact of consumption of lead via groundwater on a range of health, educational, and behavioral outcomes in the United States. See for example Aizer et al. (2018), Billings et al (2018), Trejo et al. (2021) and Zheng (2021).
[2] Other related papers analyze the link between water quality regulation and outcomes. For example, Greenstone and Hanna (2014) use a differences-in-differences setting and do not find that water regulations as being effective to reduce water contamination or infant mortality in India. On the other hand, Do et al. (2018) find that Supreme Court rulings that enforced curtailment of industrial pollution in the River Ganges led to lower incidences of infant mortality.
[3] See also: Aggarwal K. (2022), Del Razo et al. (2011), Minamoto et. al. (2005), Tseng (2007) and Watanabe et. al. (2007).
[4] See Case, Lubotsky, and Paxson (2002), Case, Fertig and Paxson (2005) and Behrman and Rosenzweig (2014) for studies in developed countries. Miguel and Kremer (2004) show that the provision of deworming drugs in Kenya led to an increase in educational attainment. Bobonis, Miguel and Puri-Sharma (2006) find a 20 percent decrease in school absenteeism rates following an iron supplementation and deworming intervention among 2–6 year-old preschoolers in the slums of Delhi, India.

nutritional disadvantage in childhood (Asadullah and Chaudhury, 2011; Vahter, 2007).[5] Second, groundwater contaminated with heavy metals causes gastroenteritis, weakness, fatigue and bone diseases increasing school absenteeism (Komarulzaman et al 2019; Hunter et al. 2014; Wasserman et al. 2004; 2007). Additionally, children are more susceptible than adults to contaminated water because of their relatively weaker immunity systems and higher proportion of water composition in their bodies compared to adults (Vahter, 2008). Furthermore, epidemiological evidence suggests that contaminants like arsenic cross the placenta and adversely impact health in utero and beyond (Rahman et al. 2009; Kile et al. 2016). Additionally, the consumption of contaminated groundwater causes well-documented visible skin and bone deformities that can impact socio-emotional development of children (Hassan et al., 2005). Finally, children living in households with other sick family members will be more likely to miss school to work or provide with care at home (Carson et. al., 2010).

Estimating the impact of groundwater contamination on educational outcomes using household or regional variation in contaminant levels is problematic for a variety of reasons. First, variation in access to safe water among households is likely to depend on socio-economic status, education levels, and thus both willingness to pay for safer water access, and health outcomes (Asadullah and Chaudhary 2011; Zhang and Xu 2016). Second, the intensity of economic activities is likely correlated with concentration levels of contaminants in groundwater. For example, in relatively high economically active areas, groundwater is overexploited and arsenic concentration higher. This is because naturally-occurring arsenic dissolves out of rock formations when groundwater levels are lower (Madajewicz et al., 2007; Akhtar et. al. 2022).

We study the effect of exposure to harmful chemicals in groundwater on education outcomes in India. First, we analyse data from a primary survey across all public schools in a heavily groundwater contaminated area in Northeast India, the state of Assam. To put it into perspective, Chetia et al. (2011) find that 76% of the water samples in Assam contained levels of heavy metals such as arsenic, iron, fluoride and manganese higher than the WHO recommended value. Coincidentally, Assam also has the highest maternal mortality rate in India (Meh et al., 2022).

Our main identification strategy relies on two government programs that provide time and cohort variation in access to safer sources of drinking water.[6] Differences-in-Differences (DD) estimates show

[5] A myriad of observational studies find a negative link between arsenic contamination in drinking water and education outcomes (see, for instance, Hassan et al. 2005, Murray and Sharmin 2015 and Wasserman et al. 2004; 2007; Rosado et.al. 2007, Jena et. al. 2020).
[6] These programs were implemented by Public Health Engineering Department (PHED) of Assam in 2008-2009. The PHED is the main government agency responsible for water supply.

that a greater number of years of exposure to contaminated drinking water (before government water supply was made available) is associated with lower Cumulative Grade Point Average (CGPA) in the prior grade attended and higher rates of grade repetition. In addition, exposure to contaminated groundwater leads to higher rates of school absenteeism and lower numeracy scores. An additional year of exposure to unsafe water increases school absenteeism by 6% and retention rates by 3.6%. Recent literature suggests that staggered treatment timings and treatment effect heterogeneity in DD estimates may be problematic (Goodman-Bacon, 2021). Thus, we estimate dynamic treatment effects and difference-in-difference estimates using the method proposed by Sun and Abraham (2021) and find similar results to the Two Way Fixed Effect estimates (TWFE).

Out of all heavy metals found in groundwaters of Assam, arsenic poses the most significant risk to humans.[7] And while the link between consuming high-levels of arsenic due to contaminated groundwater on a range of child health outcomes is well documented, lesser is known about its effects on school outcomes such as absenteeism and academic achievement. Thus, to deepen our understanding of our Assam survey results, we exploit variation in the fractions of clayey soil texture across districts within the same state as an instrument for arsenic levels in groundwater and measure its impact on school absenteeism and health outcomes for children and adults.

Instrumental Variable (IV) estimates imply an additional 8.4 % increase in monthly school absenteeism due to the average excess arsenic concentration above the WHO recommended safe limits in our sample. Beyond childhood, the IHDS-II data also allows us to confirm that, consistent with recent findings (Pitt et al, 2021), adults in areas with high levels of arsenic suffer from higher morbidity than their counterparts.

Our results are in line with the limited evidence on the effect of arsenic exposure on education outcomes. Asadullah and Chaudhury (2011) find that living closer to an arsenic contaminated well is associated with lower math test scores among school-age children in Bangladesh. However, Pitt et al. (2021) argue that proximity to arsenic contaminated sources is not exogenous to educational outcomes because households are likely to switch to surface water sources in areas where widespread information campaigns are conducted. More recently, Saing and Cannonier (2017) find that higher regional levels of arsenic in groundwater are negatively related to school enrolment in Cambodia. As aforementioned, this

[7] In the short run, overconsumption of arsenic leads to diarrhoea, vomiting, abdominal pain, muscle pain and skin rashes. Consuming arsenic for a long period of time can lead to fatal health outcomes such as kidney and heart failure, pulmonary and respiratory diseases, mental illnesses, cancer, skin-related diseases, and adverse pregnancy outcomes. For a review of studies in Bangladesh, see Yunus et al (2016).

identification strategy is problematic as arsenic levels are also likely correlated with regional differences in economic outcomes related to education such as agricultural practices, population density, and industrialisation. The most compelling identification used by Pitt et al. (2021) uses arsenic biomarkers obtained from a sample of rural households in Bangladesh to find a negative relation between genetic arsenic metabolization ability and a range of adult outcomes such as health, educational attainment, and labour productivity.

Our contribution to the literature is threefold. Our study is the first to measure the arguably causal impact of contaminated groundwater on educational outcomes beyond school attendance and at the intensive margin on outcomes such as grade retention and test scores. Second, we use quasi-experiments to measure the impact of exposure to contaminated groundwater on educational outcomes in, Assam, one of the most arsenic-contaminated areas in the world (Khurana and Sen, 2008). Third, our study highlights the importance of early interventions to improve access to healthcare and nutrition, to mitigate the effects of environmental hazards on vulnerable populations. Global warming is expected to reduce the rate at which rainwater seeps underground, increasing the level of toxic contaminants in groundwater. In the backdrop of climate change and over exploitation of groundwater, this paper makes a significant contribution to the literature on environmental health and education outcomes.[8]

The remainder of the paper is structured as follows. Section 2 describes the background and institutional context of the water supply scheme followed by data description from the school survey. Section 3 presents the differences-in-differences identification strategy and provides visual support for the identification assumptions. In section 4 we present the results from the school level analysis and conduct several robustness checks. Section 5 describes the IHDS-II data, the IV strategy and the associated results. Concluding remarks and policy implications of our analysis are presented in Section 6.

[8] While several studies have looked at the relation between pollution and education in developed countries (see, for instance, Currie et al. (2009) for the US and Lavy, Ebenstein and Roth (2014) for Israel), in the developing country context the evidence is scarce. Bharadwaj et al. (2017) compare sibling outcomes to find that exposure to air pollution in the womb has a negative impact on mathematics and language skills among fourth grade students in Chile.

2. Background and Data

2.1 Background: Groundwater Contamination and Water Sanitation Programs in Assam

India and Bangladesh have the largest population in the world exposed to arsenic-contaminated drinking water. Arsenic poisoning or so-called arsenicosis, has been linked, among others, to illnesses inducing kidney and heart failure, mental illnesses, cancer, skin-related diseases, and adverse pregnancy outcomes (Tseng 2007).

More than 70 million people in India are exposed to arsenic in groundwater that is above the WHO safety guidelines (Khurana and Sen, 2008).[9] Because children comprise nearly half of the affected population, arsenic poisoning is likely to be a main contributor to India's high child mortality rate (Asadullah and Chaudhury, 2011).

Among the states of India with high levels of groundwater contamination, Assam is one of the most severely impacted (Government of Assam, 2013). According to the 2011 Indian census, while on average 32 percent of Indians have access to safe drinking water, this figure is only 9.2 percent in Assam. This is because over 50 percent of households in this state rely on groundwater that is contaminated as their primary sources of drinking and cooking water.

Located on the south bank of the River Brahmaputra, Jorhat holds a relevant position in Assam as it serves as a major communication hub for the states in the north-east region of India.[10] Concentrations of arsenic in Jorhat's drinking water vary between 194 to 491 micrograms per litre in most of the habitations, particularly in the Titabor block, which is four to eight times higher than the safety limit of 10 micrograms recommended by the WHO (Saikia et al, 2017). In comparison to other water contaminants, fluoride concentration varies from 0.37 to 1.49 mg/L, sitting within safe limits (CGWB 2013). Jorhat district also has safe levels of iron except from scattered patches where filtering before consumption is recommended (CGWB 2013).[11] Figure 1 illustrates the geographical location of Titabor within Jorhat district.

---

[9] See Appendix Figure A.1.

[10] There are 35 states and union territories in India. Each state is further administratively divided into districts (also known as Zila). Further, these districts are categorized into sub-districts where the lowest administrative unit is a town (urban areas) or a village (rural areas). The primary sampling unit for our survey is a habitation, where a group of households collectively form habitations.

[11] The impact of the interactions between arsenic, iron, and fluoride on health are not straightforward. Mice exposed to both arsenic and fluoride had worse health outcomes such as DNA and brain damage, therefore being more likely to develop cancer and possibly cognitive delay, than those exposed to arsenic alone (Flora et al, 2009). However, the presence of iron seems to offset the effects of arsenic by absorbing it, despite it being harmful on its own at high levels (Hao et al, 2018). Given that the presence of fluoride and iron is relatively low in the region where the primary survey was conducted and their interactions with arsenic have opposite effects, we believe that the results are unlikely to be driven by one of these interactions.

While a widely used and effective tool to remove iron from drinking water are sand filters and 48% percent of households across 27 districts of Assam treat their drinking water with ceramic, sand, or other filters (National Family Health Survey, NFHS-4), none of those methods remove arsenic from groundwater.

Due to the lack of off-the shelf methods to effectively remove arsenic from groundwater, the best alternatives are rainwater harvesting when feasible, or drinking filtered and boiled surface water (Nguyen et al. 2013; Jha et al. 2017). However, group discussions conducted as part of our initial survey revealed that the high costs of maintenance rendered the implementation of rainwater harvesting and filtration techniques impractical in a predominantly rural region.

In response to these health concerns, the Assam Public Health Engineering Department (PHED) introduced two central government sponsored flagship programs, namely, the National Rural Drinking Water Program (NRDWP) and the Swachh Bharat Mission (SBM) in the Titabor region. The two schemes progressively provided access to treated and safe drinking water from two nearby rivers: Doyang and Dhansiri. The criteria for determining which habitation received safe water first was based on proximity to a Water Treatment Plant (WTP). Thus, habitations located near a WTP would be provided with safe water access first, while the construction of pipelines progressively extended further from the WTP. Figures A.2 and A.3 in the appendix provide official PHED maps of the rivers, WTPs and the geographical location of each village relative to the WTPs. Section 3 discusses in detail how the timing and geography of implementation could potentially impact our estimates.

Clean water was provided to a group of households within a village, referred to as habitations, through community taps or private connections. The National Rural Drinking Water Program (NRDWP) aimed to cover 507 habitations, benefiting approximately 40,000 individuals across 17 councils (Gram Panchayats) in the Titabor block of Jorhat district, Assam. The program offered two options for accessing safe water: households could either opt for a private connection or choose a more cost-effective shared community-level connection.

To facilitate adoption of the schemes, every habitation formed water committees of 15 representatives. Households had to sign an agreement with the president of the water committee (known as Panee Samitee) to be connected to piped water. In addition, beneficiary households have to pay set-up fee of INR 1,000 (approximately USD 14) plus INR 100 monthly for connection charges.[12] Informal

[12] The average monthly per capita income in Jorhat district is approximately INR 3,200 (Human Development Report, Assam 2014).

conversations with PHED officials revealed that the majority of households in Titabor opted for cheaper communal connections that shared costs among members. Water was provided twice a day for two hours both in private and communal connections. While the committee organised the collection of the funds to pay for communal charges, in case of any shortfall, the PHED took care of the shortfall.

2.2. Newly Collected School Survey

In 2018, we surveyed 117 primary and secondary schools in Titabor and merged this newly collected student and school data with school administrative data.[13] According to the latest available Census in 2011, the sub-district Titabor has 162 villages with a total population of 201,79, out of which approximately 90% live in rural areas.

We first conducted focus group discussions in May 2018 followed by a pilot survey. The main survey was conducted across all 3rd, 5th, and 8th across 283 habitations. According to administrative data provided by the education department, 4,316 students were enrolled across the three grades in the selected schools. We surveyed 3,065 students who were present on the first visit and an additional 446 students during a follow-up survey, which help reduce attrition from 29% to 19%.[14]

We administered both a principal and a student questionnaire. The latter captures household characteristics, whether the child faced any unfriendly atmosphere at school (if yes, then the reason), number of days absent in previous month (verified with school administrative data), and reasons for absenteeism. The questionnaire also included information on child's awareness of water contaminants, and the primary source of drinking water at home. We categorised tap water, filter/sand filters, rain water harvesting and piped water supply as safe, and hand pumps, tube well, and/or pond water as unsafe.[15] The school principal survey was designed to measure school quality characteristics such as playground, library and toilet availability, student-teacher ratio, teachers' experience, enrolments, and class size.

[13] Figure 2 depicts the details of the sampling process.

[14] School-based surveys face potential selection bias resulting from absenteeism on the day of the survey. This could underestimate the results from our study as the children who are absent on any school day are precisely the ones most likely to be affected by adverse health conditions. To minimize the bias due to absenteeism, we revisited schools on the final exam day which were scheduled in the last week of July 2018. Since schools were closing for summer vacations on the last day of exam, giving us a short time frame to conduct revisits, we revisited only those schools where the number of absent children were high on the initial date of survey. On the second visit, we surveyed only those students who were absent on the initial day of survey.

[15] We also collected information on the source of drinking water at school. However, due to constant switching of water source by school, it was difficult to trace out the actual source of drinking water while in school. For instance, there were some schools which considered water from rainwater harvesting as their primary source of drinking water. At the same time, due to the failure in its maintenance, students ultimately consumed water from hand pumps and tube wells. Further, 82% of children in our sample carried water bottles from home.

Our three main dependent variables using school administrative data are: days missed school in the past 30 days, CGPA in the previous academic year, and if the child has ever repeated a grade. In addition, we administered short grade-specific literacy and numeracy assessments mimicking the National Achievement Survey (NAS), a 2017 national survey assessing learning outcomes of Indian children, administered by the National Council of Educational Research and Training (NCERT).[16]

Table 1 provides means and standard deviations of key outcomes variables in our analysis. Just about 68 percent of the sample has access to safe drinking water at home. Only 15 percent of children could answer correctly all three math questions while 17 percent could not answer any of the questions correctly. The scores are even lower in case of verbal test (3% and 37% respectively). CGPA scored in previous grade does not appear as low, with 31 percent of students scoring above 71 percent. Table 2 provides the means of control variables such as age and gender, caste, religion, mother's education, structure of house, land/home ownership, durable assets, and ownership of heavy vehicles. As expected, the socio-economic background of our survey children is also quite poor, with 46% of their mothers being illiterate.

We obtained PHED administrative data on government water supply schemes that were constructed at the habitation level between April 2009 and March 2018. This data includes population by habitation, year in which the safe piped water scheme was installed, number of piped versus unsafe groundwater schemes, start and completion date for the project.[17] To create the main explanatory variable, we use data on year reported which is defined as the first year in which a particular pipe water scheme was reported to be installed in a habitation.

3. Empirical Strategy

3.1 Two-way Fixed Effect Differences-in-Differences

We start with a two-way fixed effect diff-in-diff (TWFEDD) approach. The identification strategy using the school-based survey exploits exogenous variation in timing, coverage of the government water supply

[16] To measure numeracy among 3rd, 5th and 8th graders, three questions about mathematical operations (addition, subtraction, multiplication, geometry and linear algebra) were assessed depending on the grade of the student. Similarly, verbal abilities were assessed for 5th and 8th graders based on three questions on reading comprehension (reading an advertisement or small passage in English). Verbal ability tests were not conducted among third graders based on the pilot revealing rather poor English language abilities among younger children. The outcome variables are the fraction of questions answered correctly by the student. The appendix lists the grade and subject specific questions that were administered.

[17] Groundwater schemes include shallow and deep tube wells. The tube wells that were constructed under these schemes were found to be contaminated due to the presence of high levels of metals particularly arsenic owing to overexploitation of groundwater for domestic and agriculture requirements. Both datasets used in this study are described in tables A.0.1 and A.0.2 in the Appendix.

scheme, and child year of birth to measure the effects of safe water access and exposure. In our estimating equation

$$\mathbf{Y_{igsh}} = \boldsymbol{\beta_1}\mathbf{years_{igsh}} + \boldsymbol{\beta_2}\mathbf{years_{igsh}} * \mathbf{Male} + \boldsymbol{\beta_3}\mathbf{X_{igsh}} + \mathbf{G_i} + \mathbf{S_i} + \mathbf{H_i} + \boldsymbol{\varepsilon}_{\mathbf{igsh}}, \quad \mathbf{(1)}$$

Y_{igsh} refers to the education outcomes for child i in grade g, school s and habitation h. $Years_{igsh}$ indicates the number of years the child has been exposed to unsafe water. This variable is determined by the interaction of age of the child, the timing of safe water availability, and habitation of residence.

For instance, in a habitation that obtained access to a safe piped water in 2013, the number of years habitation had access to safe water in 2018 is 5. If a child living in this habitation is 12, she will have been exposed 7 years to unsafe water. β_2 is the coefficient on the interaction between gender (male=1) and $Years_{igsh}$. It captures the differential effect, if any, by gender of the child based on studies that find larger health impact of groundwater contaminants among girls (Gardner et. al., 2013). X_{igsh} is a vector of individual and family and household background characteristics such as age, and dummies for gender, religion, caste, assets, and mother's education. G is a grade fixed effect, S is school fixed effect and H is habitation fixed effect. Standard errors are clustered at the habitation level.[18]

The Differences-in-Differences (DD) identification strategy exploits two sources of variation. First, we compare children in the same grade and school living in different habitations that got access to water at different points in time. Second, we compare children in different grades living in the same habitation where variation in "treatment" comes from age/cohort difference. Further, we conduct an event study type analysis and show that the treatment effect increases with years exposed to contaminated water.

One concern with the identification strategy is that parents who are aware of the water contamination problem may respond by campaigning for the water supply scheme to reach their village. In that case, we would observe that piped water is more likely to be supplied to habitations with more educated and thus more aware parents. However, survey data reveals very low levels of information on arsenic contamination among the population. Mahanta, Chowdhury and Nath (2015) find that 86% of households in arsenic affected habitations of Titabor did not know about the prevalence of arsenic in groundwater. Similarly, our survey finds that only 7.5% of children studying in public schools in Titabor were aware of arsenic groundwater contamination. As an additional robustness, we further include a control variable

[18] Note that exposure to contaminated water is defined by the interaction of age and habitation. Thus, our identification strategy cannot separately identify the effect of exposure from the effect of habitation-specific aging patterns.

for the child's awareness related to presence of arsenic, fluoride or nitrates in groundwater. Following the evidence of a consistent link between parental perceptions of the food environment and children's eating habits (Ravikumar et all, 2022), we implicitly assume that, if parents are informed about water contaminants, this should reflect in their children's awareness. Our results are robust to this inclusion.[19] While we cannot rule out the influence of elites in the timing of the rollout, we believe that the affordability of access, the community involvement in the rollout, the geographical location of villages relative to the WTP and the inclusion of a range of socio-economic controls such assets, mother's education, and caste in the regressions alleviate this concern.

Another concern with the identification assumptions is that the timing of construction might be driven by unobserved habitation level characteristics. Though we do not have data on the detailed characteristics at the habitation level, we study the relation between timing of water access and household-level characteristics that are potentially correlated with demand for environmental quality. In tables 3 and 4, the response variable is an index that is computed from the year safe water supply was made available in a habitation. For instance, if a habitation got access to safe water in year 2009 then it is given a value of 1. Similarly, if the year reported is 2018 the variable will take value nine.

Table 3 shows the coefficients on household characteristics where the outcome variable is the year index. These results confirm that none of the household characteristics are correlated with the timing of water access. Similarly, in Table 4, we also find that village level characteristics are uncorrelated with the timing of construction of water supply schemes. We also find no correlation between school quality measures aggregated at the habitation level and timing of water access and awareness of contaminants.[20]

Our estimates will also be overestimating a positive impact of safe water access if access was coincidentally being provided jointly with any other policies targeted at improving educational outcomes.[21] Between 2011 and 2012, the Assam government, distributed free bicycles to school going girls from low-income households up to grade 10 who were studying in government schools to increase

---

[19] We also explore if the effect of exposure is larger among disadvantaged groups by including in the regressions interaction terms of exposure with family background characteristics. We find that the negative effect of one more year of exposure to contaminated water on school attendance is larger among children whose mothers have lower education and those belonging to Scheduled Caste (SCs) and Scheduled Tribes (STs). There is no heterogenous effect of family background characteristics on any other measures of education. Results available upon request.

[20] These results are shown in tables A.2 and A.3 in the appendix. Note that, since the right hand side school quality measures are aggregated at the habitation level and there are several habitations in the sample that do not have any schools, the sample size shrinks drastically in these regressions.

[21] Examples of government policies implemented in the 2010s targeting children enrolled in grades 9-12 include Aarohan, Saptadhara, cash, or laptop award schemes. However, our sample is restricted to 3rd, 5th and 8th graders and were therefore directly unaffected by these schemes.

girls' enrolment rates. For this policy to bias our results, the timing of provision of bicycles would have to coincide with the timing of access to water across habitations. Though this is unlikely, we control for asset ownership including bicycles in all our regressions.[22]

A final concern with the results from the school survey is that the construction of piped water could have also changed the time cost of water collection. In a recent meta-analysis, Orgill-Mayer (2022) document a strong relation between a range of Water, Sanitation, and Hygiene (WASH) interventions and cognitive development in children. Nonetheless, the author notes that improved safe water supply access may affect educational outcomes through the time costs of water collection as well as the reduction in contaminated water consumption. In developing countries like India, women and girls are responsible for fetching water from distant areas. This might lead to significant loss of their time that they could have spent on other productive activities and in turn could result into undesirable social and economic outcomes such as lower school attendance and poor educational outcomes (Nauges and Strand 2017; Shimamura et al. 2022). But such adverse outcomes might be less prevalent in regions with availability of surface water in abundance. Assam is situated on the banks of the river Brahmaputra and its 11 tributaries and also receives more than 1500 mm of rainfall every year. Recent statistics by National Sample Survey (NSS 76th round 2018) suggest that women and girls from rural households in India spend over 30 minutes gathering water, three times more than the average 10 minutes in Assam. In our district of study, Jorhat, women and girls spend a maximum 15 minute to a minimum of 5 minutes fetching water a week. Thus, we do not expect the results to be biased due to this reason.

3.2 TWFEDD Cohort Analysis

We provide first some visual evidence in support of the identification assumptions. If drinking arsenic contaminated water for longer periods of time has adverse effects on child health and therefore on education, then, the longer a child has been exposed to arsenic contaminated water, the worse should be her/his educational outcomes. Based on this intuition, consider the following variant of Equation (1):

[22] In September 2012, the government of Assam jointly with UNICEF implemented water sanitation and hygiene policies (WASH) in schools. This program provided access to toilets particularly for girls in schools. Randomized controlled trials that assess the impact of different school-based WASH interventions find positive effects on educational attainment of students (Freeman et al., 2012; 3). Nonetheless, the timing of implementation of this scheme in Assam does not coincide with timing of access to water across habitations. We also include school fixed effects in all regressions to control for differential access to facilities under WASH across schools. Results are robust to an alternative specification in dropping school fixed effects and controlling for school quality measures including provision of toilet and availability of safe drinking water, school infrastructure (library, daily hours of electricity, playground), teacher experience, class size and student teacher ratio. Results available upon request.

$$\mathbf{Y_{ighs}} = \boldsymbol{\beta}_{\mathbf{j}} \sum_{\mathbf{j=0}}^{\mathbf{18}} \mathbf{D_{ij}} + \boldsymbol{\beta}_{\mathbf{2}} \mathbf{X_{igsh}} + \boldsymbol{\beta}_{\mathbf{3}} \mathbf{treat_h} + \mathbf{G_i} + \mathbf{S_i} + \mathbf{H_i} + \boldsymbol{\varepsilon}_{\mathbf{igsh}} \quad (\mathbf{2})$$

Where, $\mathbf{D_{ij}}$ is a dummy variable for whether student i has been exposed j number of years to unsafe water based on her birth year and the year when safe water became accessible in her habitation. Our goal is to estimate the average treatment effect of being exposed for **j** periods**.** Thus, the higher j, the longer a child has been exposed to unsafe water in her habitation. Students who have been exposed to unsafe water for up to 5 years are the omitted category[23]. $\mathrm{Treat_h}$ is a dummy variable equal to 1 if a habitation has access to safe drinking water. All other control variables mimic Equation (1) and are described in Table 2. Figure 3 plots the coefficients $\boldsymbol{\beta}_{\mathbf{j}}$ i.e. of the number of years a child has been exposed to unsafe water on school absenteeism, CGPA scored in prior grade, grade retention and numeracy. The vertical lines depict 95-percent confidence intervals around the estimates.[24]

Results are striking; all coefficients increase with years exposed with the strongest results for CGPA and grade repetition. The magnitude of the effect on school absenteeism is relatively negligible until 9 years of exposure, and starts to increase thereafter. Compared to children who have been exposed to contaminated water for less than 5 years, children with 6 years of exposure to unsafe water are approximately 10% more likely to repeat a grade while those with 10-11 years of exposure are 20% more likely to repeat a grade. Similar increasing trends are visible for numeracy scores yet coefficients are statistically significant only for children who are exposed to contaminated water for at least 13 years. While all education outcomes are worsening with an increase in the years of exposure, the estimates post 10-11 years of exposure is not monotonic for at least two of the outcomes; grade repetition and CGPA. The main reason for this is the disproportionate number of non-treated habitations as years of exposure increases. Most habitations had, by 2018, got access to clean water.[25] Thus, children who were exposed to contaminated water for more than 11 years are disproportionately more likely to belong to never-treated habitations. We conduct a robustness exercise dropping never-treated habitations in the following section.

3.3 Staggered DD Treatment and Treatment Effect Heterogeneity

[23] According to the WHO and The United States Environmental Protection Agency (EPA), the first symptoms of long term exposure to arsenic starts with skin lesions after a minimum exposure of five years. https://www.who.int/news-room/fact-sheets/detail/arsenic
[24] Some age groups have been combined due to a fall in sample size as number of years of exposure increases.
[25] We have a total of 283 habitations, out of which 70 never received clean water.

In a recent paper, Goodman-Bacon (2021) examines the TWFEDD regression model and finds that with dynamic treatment effects, even if parallel trend assumptions are met, staggered DD treatment effect estimates can yield estimates with the wrong sign. This is because the TWFEDD estimator is a weighted combination of all possible 2×2 DD estimators found in the data. Thus, treated units that are treated earlier act as control units for units that are treated later. In addition, if treatment effects vary across units or over time, the treatment can cause units to be on different trends potentially leading to biased estimates (Borusyak et al. 2021). Thus, even though we allow for our average treatment effects, β_j in Equation 2, to vary over time, the DD estimates presented in this paper may be problematic in the presence of heterogenous treatment effects and staggered treatment timings.

Sun and Abraham (2021) propose an alternative parametric specification, the Interaction-Weighted (IW) estimator, that is robust to heterogeneous treatment effects and uses the "never-treated" group as the control group. This strategy ensures that individuals that receive treatment at any point are not compared to those that previously received it. In the first step, they estimate the individual cohort-time-specific treatment effects, allowing for treatment effect heterogeneity. The second step aggregates the individual treatment effects to produce the parameter of interest which they call as the "cohort average treatment effect on the treated" or CATT.

We next estimate dynamic treatment effects using the method proposed by Sun and Abraham (2021), SA hereafter. SA focus on cohort-specific average treatment effects, CATT, in which a cohort represents all individuals receiving their first treatment at the same time. Thus, we specify a cohort as all children receiving access to clean water at the same age. Their strategy estimates the heterogeneous treatment effect of each of these cohorts, and then calculates the average of these cohort-specific estimates applying cohort-specific weights. Note that in their specification the CATT is the cohort-specific average change in outcome relative to never being treated. Since the "treatment" in our analysis is the years of exposure to contaminated water, the control group in this specification are those children who always had access to safe water. However, there are only 43 children across 8 habitations in our sample who were completely unexposed (they were born after the habitation already received clean water). SA suggest that in the setting where there is no never-treated group, one can use the last cohort (in our setting, those who had access to unsafe water for a short time) to be treated as a control group. According to the WHO and The United States Environmental Protection Agency, the first symptoms of long term exposure to arsenic starts with skin lesions after a minimum exposure of five years. Based on this, children who have less than five years of exposure to contaminated water are considered the control group for the analysis.

However, the overall results shown do not change if we limit the age to less than 3 or less than 4 years of exposure, even though the sample size for the control group drops.[26]

The SA strategy, which is robust to heterogenous treatment effects, is based on two identification assumptions, namely, parallel trends in baseline outcomes and no-anticipation effects. The parallel trend assumption is unlikely to be satisfied for the never treated units, i.e. children living in habitations that never got access to clean water. This could also explain why our dynamic DID estimates in figure 3 show slight non-monotonicity as years of exposure to contaminated water increases. There are 437 children in our sample belonging to never-treated habitations, SA recommend excluding these observations from our sample to avoid violation of the parallel trend assumption. Based on this, we remove these observations from the sample noting that results do not change even if we include them.

Violations of the no anticipation assumption in our setting would imply that parents are making some choices in periods before the initial treatment that have positive benefits for education outcomes of children. Since we only observe individuals after they have received access to clean water, we cannot directly test if there is anticipatory behaviour prior to treatment. However, such choices would be important if households were aware of the potential effects of drinking contaminated groundwater. As noted previously, survey data reveals very low levels of information on arsenic contamination among the Titabor population, thus alleviating any concerns about anticipation effects.

Figure 4 shows the event study graphs using the Sun-Abraham estimator.[27] All estimated coefficients are close to the TWFEDD model shown in Figure 3 confirming that our results are not biased by variation in treatment timing and heterogenous treatment effects. Figure 4 also further supports the dynamics of treatment effects: the Sun-Abraham estimated treatment effects are increasing with the years of exposure to treatment.[28]

4. Results and Threats to Identification

4.1 Results

[26] There were 1358 children, or 39% of the sample who were exposed to unsafe water for a maximum of five years (43 children for 0 years, 54 for 1 year only, 114 for 2 years, 291 for three years, 375 for four years and 481 for five years).

[27] The graphs show the cohort-time specific treatment effects that are computed by the *eventstudyinteract* Stata command.

[28] Recent work by Borusyak et al. (2021), Gardner (2022), and Liu et al. (2022) propose estimators that, under certain assumptions, may be more efficient than those in Callaway and Sant'Anna (2021) and Sun and Abraham (2021). These methods use all untreated observations (never treated or not-yet-treated units) to generate a potential outcome or counterfactual prediction of the outcome variable, absent treatment. What stands apart in these strategies is the existence of a robust untreated group. We decided to not use these methods as the benchmark model as we have a very small number of "untreated" observations in our data.

Next, we show results for the DD estimates corresponding to Equation 1 in Table 5. None of the interaction terms are significant and thus, there is no statistically significant difference between girls and boys in the effect of exposure. While there is no effect of years of exposure on verbal scores, an additional year of exposure to arsenic contaminated water leads to lower numeracy scores. Column 3 shows that a one-year increase in exposure to unsafe water is associated with a 0.2 day increase in absenteeism per month. This translates to approximately 6.2% increase in absenteeism with respect to the mean (3.4 days). The probability of repeating a grade increase by 3.7% with an additional one year exposure to contaminated water. An additional year exposed to contaminated drinking water also has a negative and significant (5% level) effect on CGPA scored in the previous grade (coefficient 0.09). To compare this to the effect of other environmental pollutants, Nilsson (2009) finds, among Swedish children, that when the average lead exposure during early childhood increases by 1 μg/kg, the GPA decreases by 0.017.

We find that exposure to contaminated water increases retention rates. At the same time, if retained kids underperform or are more likely to be absent, then we are capturing the effect of exposure on test scores/absenteeism via its effect on retention. To test if retained children are driving part of the treatment effect, we drop these children from the sample and re-estimate the effect of exposure. Almost 15.5% of the sample (541 students) were retained at least once. Table 6 shows that exposure continues to have a statistically significant negative effect on absenteeism and percentage scored, however the effect on math scores becomes insignificant with a smaller coefficient. Thus we can conclude that the effect of exposure on math scores is being driven by retained children.

4.2 Threats to Internal Validity and Robustness Checks

A concern with the identification strategy is a correlation between arsenic contamination and distance to the river. Wallis et. al. (2020) find elevated levels of arsenic in riverbeds suggesting a possible correlation between arsenic concentration and distance to river. If the distance to the river leads to differences in the cost of access to clean water, then, arsenic contamination would be correlated with the timing of piped water supply. This suggests that areas with higher arsenic contamination (those closer to the river) could have received piped water access earlier.

Using the official PHED pipeline installation maps and location coordinates (shown in appendix), we measure the distance in kms between the habitation from their corresponding WTP.[29] As expected, the correlation coefficient between distance from WTP and years since water access is negative (-0.30).

[29] For 57 habitations we could not ascertain the coordinates and thus were not able to compute the distance from their WTP.

Thus, as the distance from the WTP increases, the more recently has water been made available in those habitations. Since distance does not vary within a habitation, we cannot include habitation fixed effects in a regression which also includes the distance variable. Instead, we divide the sample into early treated and late treated habitations based on distance to the WTP.[30] The median distance from a WTP in our sample is 32.42 kilometres. We divide the sample into early-treated habitations (less than median distance from the WTP) and late-treated habitations (greater than or equal to median distance from the WTP). Then we estimate Equation (1) separately for the two samples. The results are shown in Table 7. The estimates for percentage scored and grade repetition are robust across the two samples and comparable to the main estimates in Table 5. Math scores have comparable coefficients but are imprecisely estimated. Absenteeism is statistically significant in Column (3) but not in Column (1), however, the coefficients are same for both late and early treated habitations. These results suggest that not only was treatment assigned based on geographical distance, but also, the effect of exposure does not differ by distance from the WTP.

As discussed previously, our survey was conducted in a region that has alarming levels of arsenic but acceptable levels of other pollutants in groundwater. While high levels of fluoride have been found in other districts of Assam, the natural levels of fluoride are low in Titabor block (Saikia et al, 2017). Further, there is a high awareness regarding the presence of iron in drinking water in Assam and households have traditionally used sand filters that effectively remove iron, but not arsenic. In our sample, while only 7.5% of students were aware of arsenic, while 51% were aware of the presence of iron. Due to the high awareness of iron presence in water, low levels of iron in most areas in Assam, and widespread use of sand filters, we do not expect the water supply schemes to have a significant effect on iron exposure levels. While we cannot directly test which groundwater contaminant led to worse education outcomes, in the next section we explore the effect of one such contaminant, namely, arsenic.

Note that while we focus on one potential contaminant, we cannot explicitly rule out the role of biological water contaminants. The construction of piped water schemes could have also led to a decline in exposure to biological contaminants. Reassuringly, the usage of sand filter is very high in Titabor (52%) and only 7% of the surveyed children were drinking surface water without filtration. This is consistent with data from the 2015-16 round of the National Family Health Survey (NFHS-4). The survey finds that 48% percent of households across 27 districts of Assam treat their drinking water with ceramic, sand, or other

[30] An alternative way to show random timing of water access is to limit the sample to never-treated and always-treated habitations and then compare outcomes across early and late treated habitations. While there are 437 respondents that were never treated , we do not have a sufficiently large sample of always treated as only 43 children were born after the habitation received water supply.

filters. Sand filters effectively removes biological contaminants, solid matter and iron from water thereby making it potable. Thus, we do not expect the results to be driven by variation in exposure to biological contaminants though we cannot rule them out.

5. Effects of Arsenic in Groundwater on School Absenteeism Using a National Survey

5.1 India Human Development Survey (IHDS)

We use district level data from the India Human Development Survey –II (2011-12). The IHDS covers around 42,152 households and 204,568 individuals across 1,503 villages and 971 urban neighbourhoods in India. The IHDS is the only nationally representative dataset in India providing information on school absenteeism with a wide range of individual, household and family background characteristics.[31]

Our main outcome variable, school absenteeism, is defined as the number of days a child was absent from school in the previous month.[32] There are 14,346 children in our sample between the ages of 6-19 enrolled in primary and secondary schools. Table 8 summarizes the key variables used in the IHDS analysis including the district level controls. On average, children missed 4.5 days of school in the previous month or approximately one school day a week. The average child in the sample is 11.8 years old and the sample is gender balanced.

5.2 Other Variables

We gather data on total area under rice and wheat production (in millions tonnes) from the Ministry of Agriculture and Farmer's Welfare and district level sex ratio, literacy rate, and urbanization from the 2011 Indian Census. We control for district level educational characteristics such as gross school enrolment and number of schools in government as well as private schools. Data for these variables is provided by District Information System for Education (DISE 2011) under National University of Educational Planning and Administration, Government of India.

Data on arsenic, fluoride, and iron levels in groundwater was obtained from the Central Ground Water Board of India. Following the WHO guidelines, the Bureau of Indian Standards (BIS) aims for a standard of 50 μgL^{-1} (microgram per litre) for arsenic in drinking water. The level of arsenic in groundwater is

[31] The ASER (Annual Status of Education Report) is an alternative annual survey that also focuses on learning outcomes and schooling status of children for rural districts in India, but it only provides information on rural areas. Because arsenic levels increase with anthropogenic activities, urban areas could potentially have high levels of arsenic in groundwater and hence the use of this dataset is not appropriate.

[32] While the IHDS provides a numeracy scores, the answer to each item is only provided if the previous item was answered generating an unreliable measure with high missing rates.

aggregated at the district level from block level data. We restrict the analysis to only those states where the presence of arsenic is measured beyond the threshold limit in at least one district. The final dataset comprises of more than 14,000 school going children across 13 arsenic affected states and 160 districts, where 41 districts are arsenic affected and 119 are non-arsenic affected districts. As shown in Table 8, the average level of arsenic is 106 $\mu gL{-1}$ (microgram per litre) across districts in India, remarkably higher than the threshold limit.

The data on soil texture is generated from Harmonised World Soil Database (HWSD) which was established in July 2008 by the Food and Agricultural Organisation (FAO) and International Institute for Applied System Analysis (IIASA). HWSD is global soil database framed within a Geographic Information System (GIS) and contains updated information on world soil resources. It provides data on various attributes of soil including texture and composition. Table 8 shows that average percentage of clayey soil across districts is approximately 29 percent. Among the district level characteristics, sex ratio is unbalanced at 930 females to every 1,000 males.

5.3 Instrumental Variable Approach

A variety of natural geochemical processes play a vital role in the release, and distribution of arsenic in groundwater. One of the important determinants of arsenic content in groundwater is the how long the water has been contained in the contaminated soil, which is related to soil permeability. Finer soils have relatively more particle density and lower porosity levels, and, as a result, their permeability level is relatively lower than loamy soil[33] which facilitates arsenic concentration in groundwater (Mc Arthur et al. 2001; Madajewicz et al. 2007).

Herath et. al (2016) find that in the Ganges–Meghna–Brahmaputra basin of India and Bangladesh, aquifers covered by finer sediments (clay) contain greater concentrations of arsenic in groundwater, whereas arsenic concentrations are significantly lower in aquifers with permeable sandy materials at the surface. Because arsenic concentration is higher in clayey relative to coarse soil, we exploit the variation in soil texture across districts within a state to instrument for groundwater arsenic contamination.

Our estimating and first stage equations, respectively, are as follows:

$$\mathbf{Y_{ids} = \delta Ars_{ds} + \gamma X_{ids} + D_{ds} + S + e_{ids}} \quad \mathbf{(3)}$$

[33] Loamy soil consists of a higher proportion of sandy and silty soil relative to clayey soil.

$$\mathbf{Ars_{ds} = \pi Soil_{ds} + \mu X_{ids} + D_{ds} + S + \epsilon_{ids}} \quad \mathbf{(4)}$$

We are interested in measuring the effect of arsenic on the education outcomes $\mathbf{Y_{ids}}$ of child i in district d of state s in Equation (3). The main explanatory variable Ars_{ds} indicates the concentration level of arsenic in groundwater. We instrument arsenic soil contamination using $Soil_{ds}$ i.e., the percentage of clayey soil in districts d. X_{ids} is a vector of controls including individual characteristics (age, age-squared and gender) and family background characteristics (parental education, log of total household consumption expenditure, religion and caste). S denotes state fixed effects. D are the district-specific controls including sex ratio, ratio of rice to wheat production, iron in groundwater, literacy rate, and urbanization rate. We also control for four district-level education variables, namely, total student enrolments and the number of schools in both private and public schools, respectively.

5.4 Threats to Identification

The identifying assumption is that clay soil fractions affect education outcomes only through the impact on the level of arsenic in groundwater. A threat to identification assumption is that soil texture impacts crop suitability and therefore, income. In particular, clayey-rich soil regions maybe more suitable for growing water intensive crops such as rice relative to wheat that requires relatively less irrigation. We control for the district-level ratio of rice to wheat production (measured in millions tonnes) in all models. We also control for district level male to female ratio in all regressions to control for the possibility that soil texture can affect economic outcomes through relative female to male employment rates (Carranza, 2014). As we show in the subsequent sections, the IV estimates do not change with inclusion of these additional control variables.

Note that while arsenic levels could also rise through increased use of fertilizers, the literature suggests that use of fertilizers does not alter the physical properties of soil (Carranza 2014). Unlike commercial crops like rice and wheat, arsenic-based pesticides are applied in specific crops such as fruit trees, potatoes, vegetables and berries which might alter some properties of superficial soil but not the subterranean soil used in our analysis.

However, there might be further threats to our identification strategy if clayey soil is correlated with other geographic or demographic characteristics that impact economic outcomes. We do not find any direct correlation between the proportion of clayey soil in a district on weather (rainfall and temperature), other contaminants such as iron, fluoride, nitrogen, nitrates, lead, phosphorus and potassium, economic and demographic factors (ratio of rice to wheat production, literacy, sex ratio, per capita education expenditure, male to female ratio in agricultural participation), conditional on state fixed effects.[34]

5.5 IV Results from the IHDS

Table 9, shows the first stage regression (Equation 4) and a positive and statistically significant relationship between arsenic and the fraction of clayey soil. The F-statistic is very high supporting soil texture as a strong instrument for arsenic levels. Table 10 reports OLS and IV estimates for absenteeism measured in number of school days missed in the last month. While OLS estimates in Column 1 are insignificant, the IV estimates suggest that arsenic has a positive and statistically significant (at the 5% level) effect on school absenteeism. The IV estimates in columns 2 and 3 show that results are not sensitive to the exclusion of geographic and economic control variables giving further credibility to the exogeneity of the instrument. In particular, the estimate in Column 3 suggests that a 1 microgram/litre increase in arsenic increases school absenteeism by 0.004 days per month. To put things in perspective, the threshold for the safe limits prescribed by the World Health Organization (WHO) of arsenic concentration is 10µg/ltr. The average arsenic concentration in our sample is 106µg/ltr (Table 8). This implies excess arsenic concentration above the safe limits of approximately 96µg/ltr which translates to an increase in monthly absenteeism by 0.380 days.[35] Against the mean absenteeism of 4.5 days (table 8), this implies an additional 8.4 % increase in monthly school absenteeism due to the average excess arsenic concentration during above the safe limits in our sample.

Medical literature suggests that the adverse effect of arsenic ingestion on school participation is larger for girls relative to boys because of social isolation due to visible arsenic-poisoning skin lesions (Saing and Cannonier, 2017; Hassan et al., 2005). Contrary to this, in Table 10, Column 4, we find a larger impact of arsenic on school absenteeism for boys than girls. This column includes an interaction term of arsenic and male. Both gender and arsenic are separately included in the regression. While there is a direct effect of arsenic on absenteeism, boys are disproportionately more likely to have a negative impact of arsenic exposure.

---

[34] A report from Central Ground Water Board (2022) shows high levels of lead in India in the following states: Telangana, Jammu & Kashmir, Jharkhand, Delhi, Haryana, Kerala, Madhya Pradesh, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal. Among these states, four states also have prevalence of arsenic, namely, Punjab, Haryana, West Bengal and Uttar Pradesh. Even after dropping these four states from the sample, the effect on absenteeism is still statistically significant. Due to space considerations we have not added the table in the paper that is available upon request to the authors.

[35] 96 times the coeffient in column 4 of table 10.

A concern with these estimates is that school absenteeism is conditional on enrolment. Children who are sick due to long term exposure to arsenic will be less likely to be enrolled in school. Similarly, boys or older children may work to provide family income support if adult members are suffering from arsenic poisoning. On the other hand, girls might be less likely to get treatment than boys, and hence, more likely to drop out of school. Therefore, the results could be driven by selection due to differential school enrolment rates by gender and age. To address this concern, we restrict the sample to a younger cohort (6-13 years) of children whose school enrolment is mandatory by law and below the legal child labour age of 14 in India.[36]

Results in Column 5 show that, while the magnitude of the estimate for the younger sample is larger, it is not significantly different from the overall sample and the gender interaction term becomes insignificant. Nonetheless, higher absenteeism among boys could still be explained by them missing school to work despite being enrolled. Alternatively, it could also be due to higher consumption of contaminated water and/or food compared to girls.

While we are unable to test these two channels directly as questions on hours of work in the IHDS are only asked to adults and there is no information on water and food consumption, Table 11 explores IV estimates for the health outcomes for adults and children separately. Column 1 shows no effects of arsenic consumption on short term morbidity incidence of diarrhoea among all age groups in Column 1. However, as shown in Column 2, among older children (14 to 19 year olds) a 1 milligram per litre increase in arsenic in a district leads to 8.5% increase in the probability of a diarrhoea episode in the last 30 days. Further, there is no evidence of a differential effect by gender suggesting that boys and girls are equally likely to have adverse short term incidences of diarrhoea due to arsenic consumption. In columns 3 to 5 we show detrimental effects of arsenic consumption among adults (above 19 years of age). Higher levels of arsenic in a district is associated with increasing incidence of long term morbidity among adults, consistent with epidemiological evidence. Arsenic is associated with higher incidences of diabetes and respiratory difficulties. Further adults are 6.5% more likely to have taken treatment for short term morbidity in the last one year. This could potentially explain the larger coefficient of school absences among older boys. They are not only more likely to be absent due to illnesses associated with drinking contaminated water, but also have to work to support family members who are ill due to arsenic induced diseases.

[36] 97.5% of IHDS sample is enrolled in school in this age group as opposed to 86% in the sample of 6 to 19 year olds.

To sum up, while arsenic consumption leads to higher absenteeism among children, once we account for enrollment differences, we do not find higher social ostracization induced absenteeism among girls, consistent with the results from the Assam survey. However, we find evidence of higher arsenic-induced absenteeism among older boys (14 and above) relative to girls which could be explained by an increase in work days at the expense of school attendance.

To put things in perspective, a 1 milligram per litre increase in arsenic in a district increases the incidence of diarrhoeal diseases by 8.5% among older aged school going children. Further, there is an additional 8.4 % increase in monthly school absenteeism due to the average excess arsenic concentration above the safe limits in our sample. At this point, it is worth comparing these estimates to those found in the literature on the adverse effects of air pollution in developing countries. Singh (2022) shows that an increase of 1 standard deviation (SD) in PM2.5 concentration leads to 8.3% of mean enrollment being absent on a given day. Chen et. al. (2018) find that a one standard deviation of daily AQI levels in China increases respiratory illnesses by 10.04% and the total absence rate by 7.01%.

Finally, we conduct a falsification test where we show the reduced form relation between the instrument (clayey soil) and education outcomes. If our results were driven by differences in agricultural patterns including crop type and irrigation availability or differences in female labour force participation rates, we should find a significant effect of soil type on education outcomes even in non-arsenic areas. On the other hand, if clayey soil is indeed exogenous and does not affect education outcomes through any other channels, then in areas where there is no evidence of arsenic in groundwater, clayey soil should have no effect on health outcomes. For this regression, we show the reduced form relation between clay soil and school absenteeism in districts with and without arsenic (within the 9 arsenic states). The exclusion restriction implies that this falsification test should have no effect on education outcomes in non-arsenic districts. The results in Table 12 confirm that clay soil affects absenteeism only in districts with arsenic. There is no statistically significant effect of soil type on the number of days a child is absent in districts where no arsenic is found in groundwater. Note that the coefficient for non-arsenic districts is non-trivial (almost half the coefficient in arsenic districts), however, the t-test for differences across the coefficients in the two columns is significant. Though the IV exclusion restriction is untestable, and the IV results need to be interpreted with caution, the reduced form results gives us further confidence that our identification assumption is met and clay soil does not affect education outcomes through other channels.

6. Discussion

The leading cause of morbidity in India is the lack of access to safe drinking water affecting more than 75.8 million people (WHO 2009). One study concluded that each year India loses 30.5 million 'disability-adjusted life- years' because of poor water quality, sanitation and hygiene (Brandon and Hommann, 1995). Moreover, over exploitation of groundwater and climate change is steadily increasing the concentration of toxic metals in groundwater and thus increasingly adding to the costs (Ahktar et al, 2022; Mahanta et al. 2015). These figures do not include the costs associated with the loss in human capital and its intergenerational effects. Our paper contributes to the literature measuring the effects of environmental contamination on human capital.

Combining results from a large nationally representative household survey with a primary survey conducted across schools in one of the most arsenic affected regions of India, we study the effect of contamination in groundwater on education outcomes. First, we study the education outcomes more closely in one of the most arsenic contaminated districts of India where a safe alternative to groundwater was made available via a government water supply project. We find that children exposed to unsafe water for a longer period of time had higher school absenteeism, grade retention, and lower test scores. Second, using a nationally representative dataset, we show that levels of arsenic above recommended limits have a direct causal association with absenteeism among children. Moreover, adults in areas with high levels of arsenic suffer from respiratory illnesses and Diabetes and are more likely to have taken medical treatment in the last 12 months. We also find higher diarrhoea incidences among older children. This supports the evidence that arsenic impacts education outcomes via its effect on child health, in-utero exposure, maternal health and parents labour supply.

Mahanta et. al. (2015) estimates that the annual health costs in Assam of a one microgram increase in arsenic concentration per litre to be equivalent to USD 0.01 million. The welfare gains from reducing the level of arsenic concentration to the safe limits is estimated to be USD 2.49 million. In addition to health costs, our study finds that the negative impact of groundwater pollutants on education outcomes would imply substantial economic costs associated with decreased productivity and wages.

References

Aggarwal, K. and Barua, R. (2022). Gender Disparities in the Prevalence of Undernutrition in India: The Unexplored Effects of Drinking Contaminated Water. Working Paper.

Malik Muhammad, A., Tang, Z., & Xiao, T. (2022). Evaluation of the factors affecting arsenic distribution using geospatial analysis techniques in Dongting Plain, China. *Frontiers in Environmental Science*, *10*, 1024220.

Aggarwal, K. (2023). Effect of Arsenic on Health: Evidence from Indian Districts. *The Indian Economic Journal*, *71*(2), 420-433.

Aizer, A., Currie, J., Simon, P., and Vivier, P. (2018). Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics*, 10(1), 307-41.

Asadullah, M. N., Chaudhury N. (2011). Poisoning the mind: Arsenic contamination of drinking water wells and children's educational achievement in rural Bangladesh, *Economics of Education Review* 30 (5):873-888.

Behrman, J. R., and Rosenzweig, M. R. (2004). Returns to Birthweight, *The Review of Economics and Statistics*, 86(2), 586-601.

Bharadwaj, P., Matthew Gibson, Joshua Graff Zivin and Christopher Neilson (2017). Gray Matters: Fetal Pollution Exposure and Human Capital Formation, *Journal of the Association of Environmental and Resource Economists*, 4(2): 505-542.

Billings, S. B., & Schnepel, K. T. (2018). Life after lead: Effects of early interventions for children exposed to lead. *American Economic Journal: Applied Economics*, *10*(3), 315-344.

Bobonis, G. J., Miguel, E., and Puri-Sharma, C. (2006). Anaemia and school participation, *Journal of Human Resource*s, 41(4): 692-721.

Borusyak, K., Jaravel, X., & Spiess, J. (2021) Revisiting event study designs: Robust and efficient estimation, arXiv:2108.12419.

Brainerd, E., and Menon, N. (2014). Seasonal effects of water quality: The hidden costs of the Green Revolution to infant and child health in India, *Journal of Development Economics*, 107, 49-64.

Brandon Carter and Hommann Kirsten. (1995). The cost of Inaction: Valuing the Economy-wide Cost of Environmental Degradation in India. World Bank.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225, 200–230

Carranza, E. (2014). Soil endowments, female labour force participation and the demographic deficit of women in India, *American Economic Journal: Applied Economics*, 6(4): 197-225.

Carson, R. T., Koundouri, P., and Nauges, C. (2010). Arsenic mitigation in Bangladesh: a household labor market approach, *American Journal of Agricultural Economics,* 93(2): 407-414.

Case**,** A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient, *American Economic Review*, 92(5): 1308-1334.

Case, A., Fertig, A., and Paxson, C. (2005). The lasting impact of childhood health and circumstance, *Journal of Health Economics*, 24(2): 365-389.

Cattan, S.D., Kamhöfer A., Karlsson M., Nilsson, T. (2023) The Long-Term Effects of Student Absence: Evidence from Sweden, *The Economic Journal* 133 (650): 888–903.

Central Ground Water Board. (2022). National Compilation on Dynamic Ground Water Resources of India 2022. Ministry of Water Resources, River Development and Ganga Rejuvenation, Government of India.

Central Ground Water Board. Ground Water Information Booklet (2013) Jorhat District, Assam. Jorhat: Ministry of Water Resources, North Eastern Region.

Census of India (2011). Registrar General of India, Ministry of Home Affairs, Government of India.

Chen, Siyu, Chongshan Guo, and Xinfei Huang. (2018) Air pollution, student health, and school absences: Evidence from China. *Journal of Environmental Economics and Management*, 92:465–97.

Chetia, M., Chatterjee, S., Banerjee, S., Nath, M. J., Singh, L., Srivastava, R. B., & Sarma, H. P. (2011). Groundwater arsenic contamination in Brahmaputra river basin: a water quality assessment in Golaghat (Assam), India. *Environmental Monitoring and Assessment*, *173*, 371-385.

Currie, J., Hanushek, E.A., Kahn, E.M., Neidell, M. and Rivkin, S.G. (2009). Does pollution increases school absences? *Review of Economics and Statistics,* 91(4): 682-94.

Del Razo, LM., Gonzalo G García-Vargas, Olga L Valenzuela, Erika Hernández Castellanos, Luz C Sánchez-Peña, Jenna M Currier, Zuzana Drobná, Dana Loomis, Miroslav Stýblo (2011). Exposure to arsenic in drinking water is associated with increased prevalence of diabetes: a cross-sectional study in the Zimapán and Lagunera regions in Mexico, *Environment Health*, 10(1): 73.

Do, Q. T., 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.

Freeman M, Greene L, Dreibelbis R, Saboori S, Muga R, Brumback B, Rheingans R. (2012). Assessing the impact of a school-based water treatment, hygiene, and sanitation program on pupil absence in Nyanza Province, Kenya: A cluster-randomized trial. *Tropical Medicine & International Health*. 17(3):380–391.

Fledderjohann, J., Agrawal, S., Vellakkal, S., Basu, S., Campbell, O., Doyle, P. Stuckler, D. (2014). Do girls have a nutritional disadvantage compared with boys? Statistical models of breastfeeding and food consumption inequalities among Indian siblings, *PloS one*, 9(9), e107172.

Flora, S. J. S., Mittal, M., & Mishra, D. (2009). Co-exposure to arsenic and fluoride on oxidative stress, glutathione linked enzymes, biogenic amines and DNA damage in mouse brain. *Journal of the Neurological Sciences*, *285*(1-2), 198-205.

Gardner RM, Kippler M, Tofail F, Bottai M, Hamadani J, Grandér M, Nermell B, Palm B, Rasmussen KM, Vahter M. (2013). Environmental exposure to metals and children's growth to age 5 years: a prospective cohort study. *American Journal of Epidemiology*. 2013 Jun 15;177(12):1356-67.

Gardner, J. (2022). Two-stage differences in differences. eprint: 2207.05943.

Garn, J. V., Greene, L. E., Dreibelbis, R., Saboori, S., Rheingans, R. D., & Freeman, M. C. (2013). A cluster-randomized trial assessing the impact of school water, sanitation and hygiene improvements on pupil enrolment and gender parity in enrolment. *Journal of Water, Sanitation and Hygiene for Development*, *3*(4), 592-601.

Greenstone, M., and Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in India, *American Economic Review*, 104(10), 3038-72.

Goodman-Bacon, Andrew. (2021). Difference-in-Differences with Variation in Treatment Timing *Journal of Econometrics*, 225 (2), 254-277.

Gorman, E., Harmon C.P., Mendolia, S., Staneva, A., and Walker, I. (2021). The Causal Effects of Adolescent School Bullying Victimisation on Later Life Outcomes, *Oxford Bulletin of Economics and Statistics* (84) 4.

Goyari, P. (2008). Scarcity in the midst of plenty: irrigation development for water abundant Assam. Managing water in the face of growing scarcity, inequity and declining returns: Exploring fresh approaches, IWMI-TATA Water Policy Program, April, 2, 586-600.

Government of Assam (2013). Annual report, Public Health Engineering Department, Assam.

Government of Assam (2014). Assam Human Development Report, Social Welfare Department, Govt. of Assam.

Hassan, M. M. et al. (2005). Social implications of arsenic poisoning in Bangladesh, Social Science and Medicine, 61(10): 2201-2211.

Hao, L., Liu, M., Wang, N., & Li, G. (2018). A critical review on arsenic removal from water using iron-based adsorbents. RSC advances, 8(69), 39545-39560.

Herath, I., Vithanage, M., Bundschuh, J. et al. (2016). Natural Arsenic in Global Groundwaters: Distribution and Geochemical Triggers for Mobilization, *Current Pollution Reports* 2, 68–89.

Hunter, P. R., Risebro, H., Yen, M., Lefebvre, H., Lo, C., Hartemann, P., and Jaquenoud, F. (2014). Impact of the provision of safe drinking water on school absence rates in Cambodia: a quasi-experimental study, *PloS one*, 9(3), e91847.

IIPS, I. (2017). National family health survey (NFHS-4), 2015–16. International Institute for Population Sciences (IIPS), Mumbai, India, 791-846.

Jayachandran, Seema and Ilyana Kuziemko. (2011). Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India, *The Quarterly Journal of Economics*, 126(3): 1485–1538.

Jena, A. K., Barman, M., Gupta, S., Devi, J., Bhattacharjee, S., and Barbhuiya, S. Y. (2020). Effect of Arsenic Contaminated Ground Water on the Developmental Phenomena of Primary-School Children in Assam, *Ecology, Environment, and Conservation* 26(3): 1212-1227.

Jha, S. K., Mishra, V. K., Damodaran, T., Sharma, D. K., and Kumar, P. (2017). Arsenic in the groundwater: Occurrence, toxicological activities, and remedies, *Journal of Environmental Science and Health*, Part C, 35(2), 84-103.

Khurana, I., Sen, R. (2008). Drinking water quality in rural India: Issues and approaches, Water Aid India, 2008, (288701), 31.

Kile, M. L., Cardenas, A., Rodrigues, E., Mazumdar, M., Dobson, C., Golam, M., Christiani, D. C. (2016). Estimating Effects of Arsenic Exposure During Pregnancy on Perinatal Outcomes in a Bangladeshi Cohort, *Epidemiology*, 27(2): 173–181.

Komarulzaman, A., de Jong, E., and Smits, J. (2019). Effects of water and health on primary school enrolment and absenteeism in Indonesia, *Journal of Water and Health*, 17(4), 633-646.

Lavy, Victor, Avraham Ebenstein and Sefi Roth (2014). The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation NBER Working Papers 20648.

Liu, L., Wang, Y., Xu, Y. (2022). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*, 68(1), 160-176.

McArthur, J. M., Ravenscroft, P., Safiulla, S., and Thirlwall, M. F. (2001). Arsenic in groundwater: testing pollution mechanisms for sedimentary aquifers in Bangladesh, *Water Resources Research,* 37(1): 109-117.

Madajewicz, M. et al. (2007). Can information alone change behaviour? Response to arsenic contamination of groundwater in Bangladesh, *Journal of Development Economics*, 84(2): 731-754.

Mahanta R, Chowdhury J, Nath HK (2015) Health costs of arsenic contamination of drinking water in Assam, India, *Economic Analysis and Policy* 49:30–42.

Meh C., Sharma A., Ram U., Fadel S., Correa N., Snelgrove J.W., Shah P., Begum R., Shah M., Hana T., Fu S.H., Raveendran L., Mishra B., Jha P. (2022). Trends in maternal mortality in India over two decades in nationally representative surveys. *BJOG*. 2022 Mar;129(4):550-561.

Miguel, E., and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities, Econometrica, 72(1): 159-217.

Minamoto, K., Mascie-Taylor, C. G., Moji, K., Karim, E., and Rahman, M. (2005). Arsenic-contaminated water and extent of acute childhood malnutrition (wasting) in rural Bangladesh, Environmental sciences: An *International Journal of Environmental Physiology and Toxicology*, 12(5), 283.

Motiram, S and Osberg, L. (2010). Gender Inequalities in Tasks and Instruction Opportunities within Indian Families, *Feminist Economics*, 16(3): 141-167.

Murray, M. P., & Sharmin, R. (2015). Groundwater arsenic and education attainment in Bangladesh. *Journal of Health, Population and Nutrition*, *33*(1), 1-10.

Nauges, C., and Strand, J. (2017). Water hauling and girls' school attendance: some new evidence from Ghana, Environmental and Resource Economics, 66(1), 65-88.

Nguyen, D. C., Dao, A. D., Kim, T. I., and Han, M. (2013). A sustainability assessment of the rainwater harvesting system for drinking water supply: a case study of Cukhe Village, Hanoi, Vietnam, *Environmental Engineering Research*, 18(2), 109-114.

Orgill-Meyer, J. (2022). The Evidence Base for Cognitive, Nutrition, and Other Benefits From Water, Sanitation, and Hygiene Interventions, In Oxford Research Encyclopaedia of Global Public Health.

Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2021). Identifying the costs of a public health success: Arsenic well water contamination and productivity in Bangladesh. *The Review of Economic Studies*, *88*(5), 2479-2526.

Pells, K., Ogando Portela, M.J, Espinoza Revollo, P. (2016). Experiences of Peer Bullying among Adolescents and Associated Effects on Young Adult Outcomes: Longitudinal Evidence from Ethiopia, India, Peru, and Viet Nam, Office of Research-Innocenti. UNICEF. Discussion Paper 2016-03.

Rahman, A., Vahter, M., Smith, A. H., Nermell, B., Yunus, M., El Arifeen, S., and Ekström, E. C. (2009). Arsenic exposure during pregnancy and size at birth: a prospective cohort study in Bangladesh, *American Journal of Epidemiology*, 169(3), 304-312.

Ravikumar, D., Spyreli, E., Woodside, J., McKinley, M., & Kelly, C. (2022). Parental perceptions of the food environment and their influence on food decisions among low-income families: a rapid review of qualitative evidence, *BMC Public Health*, 22(1), 1-16.

Rosado, J. L., Ronquillo, D., Kordas, K., Rojas, O., Alatorre, J., Lopez, P., ... & Stoltzfus, R. J. (2007). Arsenic exposure and cognitive performance in Mexican schoolchildren. *Environmental health perspectives*, *115*(9), 1371-1375.

Saikia, Prasanta and Kotoky, Probhat and Goswami, Uttam. (2017). Distribution of Fluoride and Arsenic Contents in the Ground Water System, Jorhat District, Assam, India, *International Journal of Science and Research* (IJSR). 6. 6-391.

Saing, C. H., and Cannonier, C. (2017). Arsenic Exposure and School Participation in Cambodia, SSRN Working Paper 2907461.

Shimamura, Y., Shimizutani, S., Taguchi, S., & Yamada, H. (2022). The Impact of Better Access to Improved Water Sources on Health, Schooling, and Water Collection of Girls and Boys in Rural Zambia. *The Journal of Development Studies*, *58*(9), 1750-1771.

Singh, T. P. (2020). Beyond the haze: Air pollution and student absenteeism—evidence from India. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3680588

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175-199.

Trejo, S., Yeomans-Maldonado, G., and Jacob, B. (2021). The Psychosocial Effects of the Flint Water Crisis on School-Age Children, WP No. w29341. National Bureau of Economic Research.

Tseng CH. (2007). Metabolism of inorganic arsenic and non-cancerous health hazards associated with chronic exposure in humans, *Journal of Environmental Biology*, 28(2): 349–357.

UNICEF. (2006). Progress for children: A report card on water and sanitation Number 5, September 2006.

Vahter, M. E. (2007). Interactions between arsenic-induced toxicity and nutrition in early life. *The Journal of nutrition*, *137*(12), 2798-2804.

Vahter, M. (2008). Health effects of early life exposure to arsenic. *Basic & clinical pharmacology & toxicology*, *102*(2), 204-211.

Wallis, I., Prommer, H., Berg, M., Siade, A. J., Sun, J., & Kipfer, R. (2020). The river–groundwater interface as a hotspot for arsenic release. *Nature Geoscience*, *13*(4), 288-295.

Wasserman, G. A. et al. (2004). Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh, *Environmental Health Perspectives*, 112(13): 1329-1333

Wasserman, G. A. et al. (2007). Water arsenic exposure and intellectual function in 6-year-old children in Araihazar, *Bangladesh, Environmental Health Perspectives*, 115(2), 285-289

Watanabe, C., Matsui, T., Inaoka, T., Kadono, T., Miyazaki, K., Bae, M. J., and Mozammel Haque Bokul, A. T. M. (2007). Dermatological and nutritional/growth effects among children living in arsenic-contaminated communities in rural Bangladesh, *Journal of Environmental Science and Health,* part a, 42(12), 1835-1841.

World Health Organization. (2009). Boron in drinking-water: Background document for development of WHO Guidelines for Drinking-water Quality (No. WHO/HSE/WSH/09.01/2). Geneva: World Health Organization.

WHO, U. (2015). WHO/UNICEF joint monitoring programme for water supply and sanitation. Estimates on the use of water sources and sanitation facilities.

UN Inter-agency Group for Child Mortality Estimation (2013) Child mortality estimates. Available: http://www.childmortality.org/.

Yunus, F. M., Khan, S., Chowdhury, P., Milton, A. H., Hussain, S., & Rahman, M. (2016). A review of groundwater arsenic contamination in Bangladesh: the millennium development goal era and beyond. *International Journal of Environmental Research and Public Health*, *13*(2), 215.

Zhang, J., & Xu, L. C. (2016). The long-run effects of treated water on education: The rural drinking water program in China, *Journal of Development Economics,* 122, 1-15.

Zheng, J. (2021). Lead exposure, human capital formation, and inequality: the impacts of drinking water lead on short-run and long-run educational outcomes, Working Paper.

Figure 1: Geographical location of Titabor Block in Jorhat District of Assam

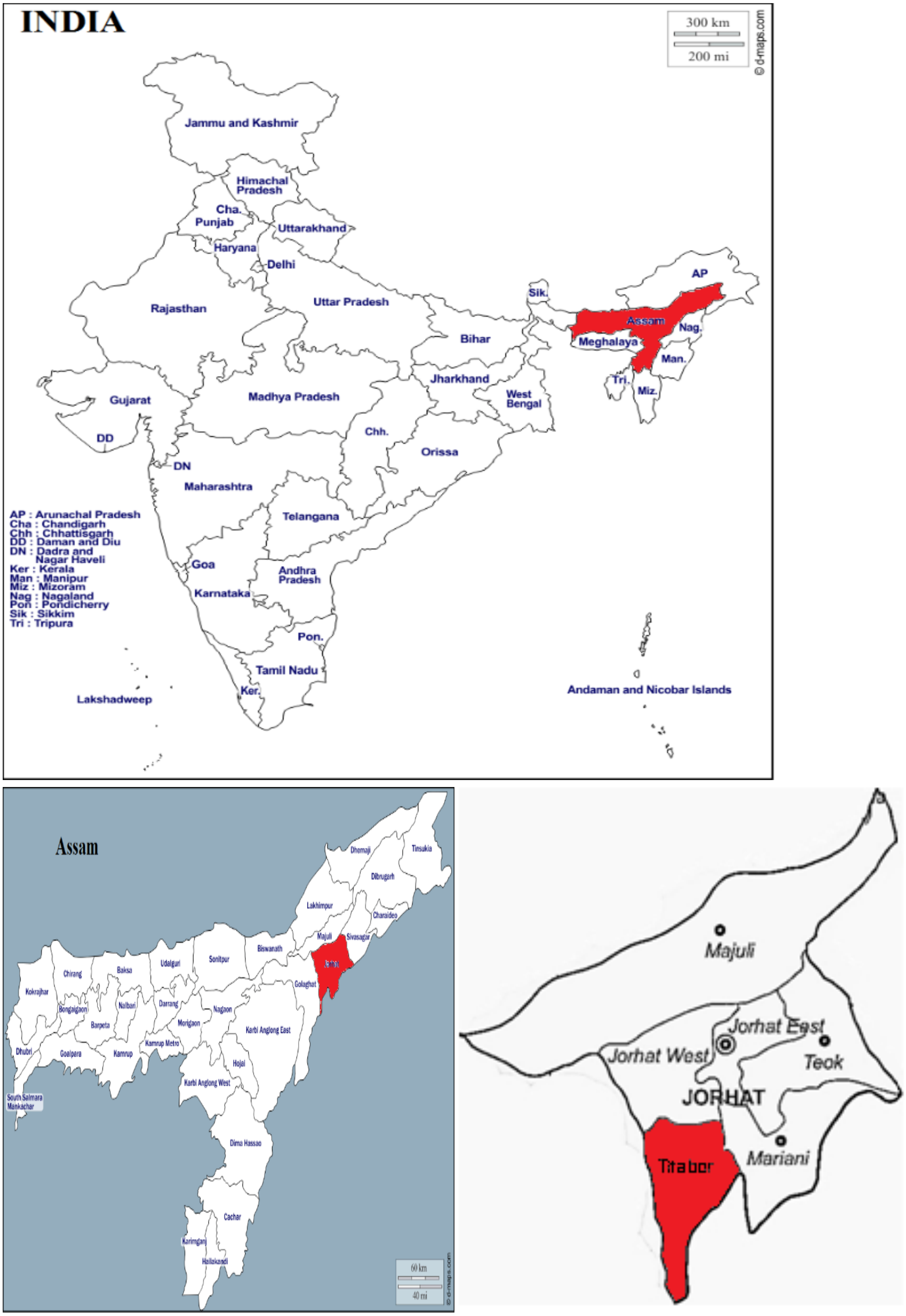
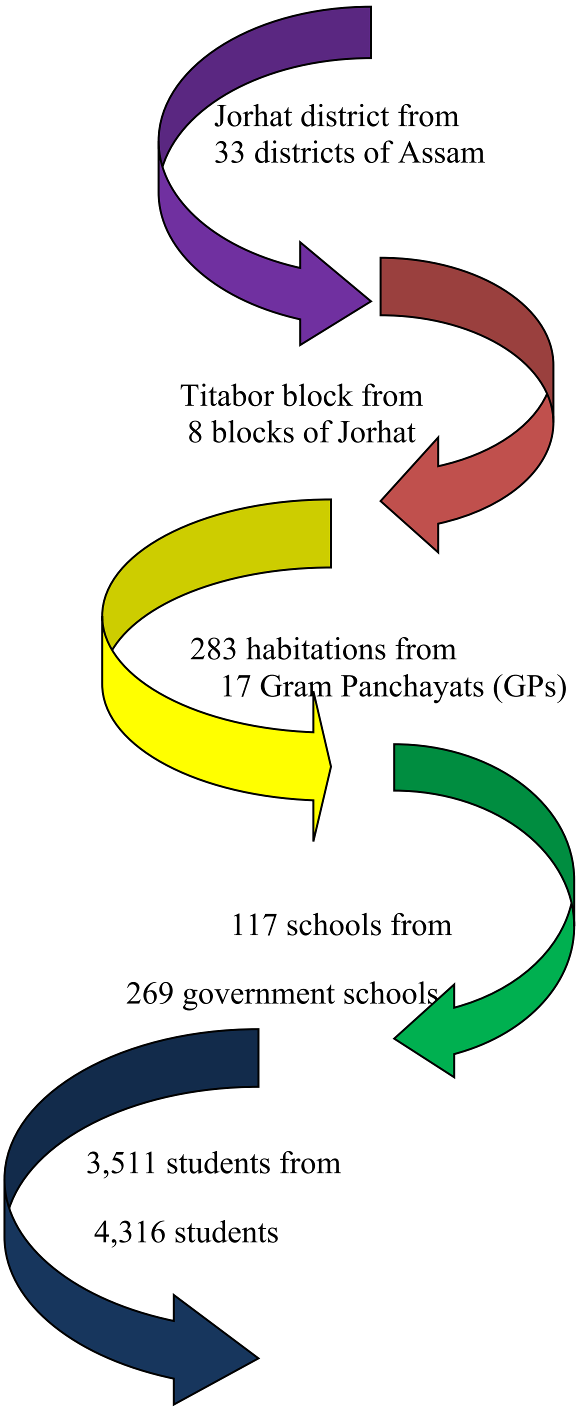


Figure 2: School Sampling Cascade



- Among the 13 arsenic affected states of India, Assam is one of the most severely impacted.

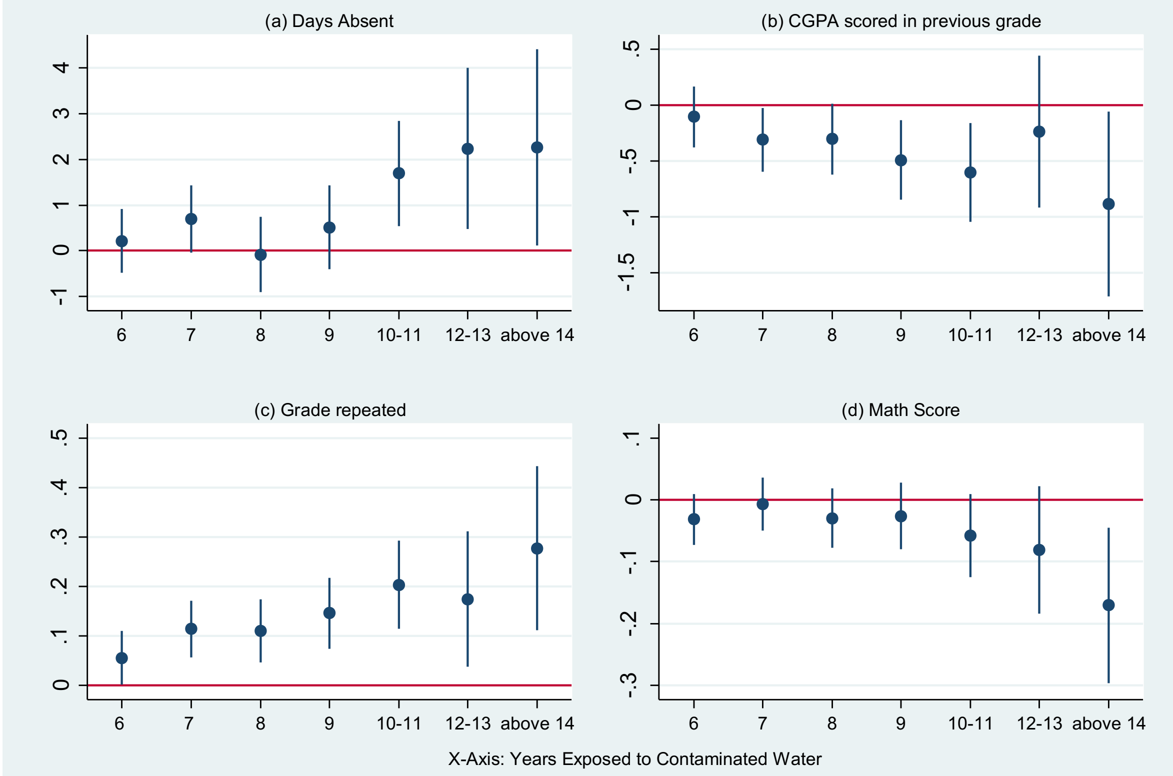
- Out of the 27 districts in Assam, Jorhat has the largest amount of contaminated habitations, namely 815 out of total 963.

- Amongst the 8 arsenic affected blocks of Jorhat, government schools in Titabor were surveyed. This selection of block was based on implementation of water supply project by PHED, Gov. of Assam.

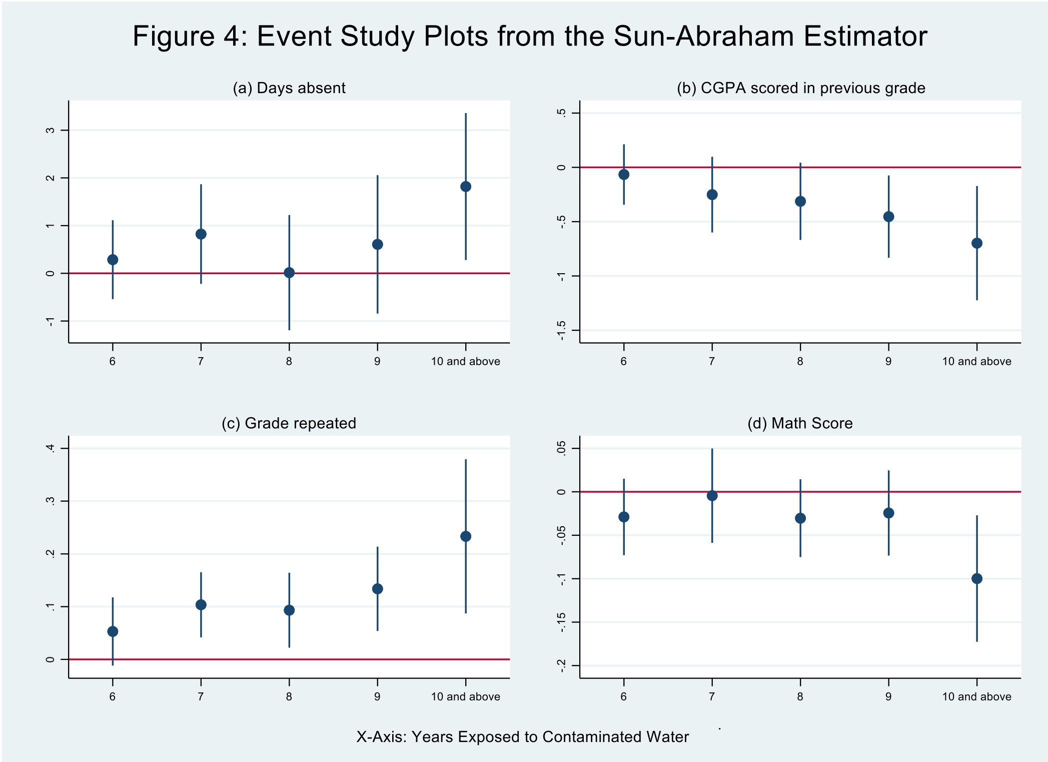
- Out of 269 government lower primary, primary, upper primary, higher secondary, and senior higher secondary, 117 schools were surveyed. We surveyed 3rd, 5th and 8th grade students. School selection criteria were based on the total number of student enrolled.

- As per the administrative data, 4,316 students were enrolled across three grades in surveyed schools. We were able to survey 3,511 students.

Figure 3: Event Study Plots from the Dynamic DID Model



Note: Figure 3 plots the coefficients $\boldsymbol{\beta_j}$ i.e. of the number of years a child has been exposed to unsafe water on educational outcomes as specified in Equation 2. Panel a (top left) represents school absenteeism, Panel b (top right) represents Cumulative Grade Point Average (CGPA) scored in prior grade, Panel c (bottom left) represents grade retention and Panel d (bottom right) represents numeracy. The vertical lines depict 95-percent confidence intervals around the estimates. Students who have been exposed to unsafe water for up to 5 years are the omitted category. All other control variables are as in Equation (2) and described in Table 2. Due to small sample sizes, years 6-7 and 10-11, 12-13 and above 14 have been grouped together as shown on the X-axis.



Note: Figure 4 plots the coefficients $\boldsymbol{\beta_j}$ i.e. the cohort-specific average change in education outcomes relative to those children who had limited access to unsafe water (less than five years) using the approach outlined by Sun-Abraham (2021). Panel A (top left) represents school absenteeism, Panel B (top right) represents Cumulative Grade Point Average (CGPA) scored in prior grade, Panel C (bottom left) represents grade retention and Panel D (bottom right) represents numeracy. The vertical lines depict 95-percent confidence intervals around the estimates. All other control variables are as in Equation (2) and described in Table 2. Children born in never treated habitations are dropped from the analysis. Due to the small sample sizes for older cohorts, ages 10 and above are clubbed together.

Table 1: Descriptive Statistics of Self-collected School Survey in Assam: Main Variables

	Mean	Std. Dev.	N
Outcome Variables			
No. of days absent in last 30 days	3.4	4.69	3,500
Ever repeated a grade (=1 if yes)	0.15	0.36	3,500
CGPA (% Scored in previous grade)			
20 to 40 %	0.10	0.3	3,387
41to70 %	0.58	0.49	3,387
>71 %	0.31	0.46	3,387
Literacy			
None of questions answered correctly	0.37	0.01	948
One questions answered correctly	0.39	0.01	989
Two questions answered correctly	0.21	0.01	529
All questions answered correctly	0.03	0.01	85
Numeracy			
None of the questions answered correctly	0.17	0.01	594
One questions answered correctly	0.33	0.01	1,161
Two questions answered correctly	0.35	0.01	1,214
All questions answered correctly	0.15	0.01	529
Main Explanatory Variable			
Safe drinking water at home	0.68	0.47	3,500
Years exposed to unsafe water	6.68	2.97	3,500

Table 2: Descriptive Statistics of Self-collected School Survey: Control Variables

Control variables	**Mean**	**Std. Dev.**	**N**
Individual characteristics			
Age	11.37	2.2	3,500
Male	0.49	0.5	3,500
Parental characteristics:			
Religion (Hindu)	0.89	0.31	3,499
Solid house structure (Brick, concrete, stone, timber and cement)	0.27	0.44	3,500
Caste			
General/Brahmins	0.18	0.01	594
Other Backward Class	0.67	0.01	2,228
Scheduled Caste/Scheduled Tribe	0.16	0.01	518
Assets:			
Land/house	0.98	0.13	3,500
Durable Assets	0.87	0.33	3,500
Heavy vehicles	0.92	0.27	3,500
Mothers' education:			
Illiterate	0.46	0.01	1,611
Primary	0.14	0.01	484
Secondary	0.33	0.01	1,158
University	0.07	0.01	247

The school level survey was conducted in 2018 among 117 government schools from the Titabar Block of Jorhat district in Assam. The questions were administered to 3,501 children enrolled in grade 3rd, 5th and 8th. The table provides the mean of control variables included in equations (1) and (2).

Table 3: Correlation between Household Characteristics and Timing of Access to Piped Water

Household Characteristic	Year of access
Other Backward Class	0.003
	(0.039)
Hindu	-0.066
	(0.054)
Solid house structure	0.060
	(0.040)
Mother's education	0.022
	(0.022)
Ownership of land/house	-0.078
	(0.149)
Ownership of durable assets	0.024
	(0.038)
Ownership of heavy vehicles	0.009
	(0.045)
Usage of sand filters	-0.006
	(0.037)
Awareness of contaminants	0.046
	(0.029)
Distance to Water Treatment Plant	-0.000
	(0.001)

N=3501. Robust SE clustered at the habitation level. *** p<0.01, ** p<0.05, * p<0.1. Coefficients on household characteristics from a regression of year of water access in a habitation on that characteristic. The index for year of access takes the value from 1 (for 2009) to 9 (for 2018). School and grade fixed effects are included.

Table 4: Correlation Between Timing of Access to Water Supply and Village Level Characteristics

Census Village Characteristics	Year of access
Total population/# of Households	0.002
	(0.258)
Children population (0-6 years)	0.001
	(0.002)
Scheduled Caste	0.000
	(0.001)
Scheduled Tribe	0.000
	(0.000)
Literacy	-0.003
	(0.007)
Total workers	-0.001
	(0.001)
Marginal worker	0.000
	(0.001)
Observations	158

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is an index for the year in which a habitation got access to safe water. The index for year of access takes the value from 1 (for 2009) to 9 (for 2018).

Table 5: Years of Exposure to Contaminated Drinking Water and Educational Outcomes: Heterogenous Effects by Gender

Variables	(1) Math Scores	(2) Verbal Scores	(3) Absent Days	(4) Grade Repeated	(5) CGPA
Years Exposed	-0.012**	-0.001	0.218*	0.037***	-0.086**
	(0.005)	(0.007)	(0.116)	(0.009)	(0.036)
Male	0.011	-0.012	0.248	0.012	-0.005
	(0.017)	(0.019)	(0.247)	(0.018)	(0.110)
Exposure*Male	0.009	0.005	-0.027	-0.002	- 0.064
	(0.007)	(0.008)	(0.099)	(0.007)	(0.048)
Observations	3,338	2,446	3,340	3,340	3,231

Robust standard errors clustered at the habitation level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All specifications include habitation, grade and school fixed effects. The main explanatory variable is years of exposure to contaminated water as shown in equation 1. The other controls are age, gender, religion, caste, assets, mother's education. The omitted categories of the control dummies are Brahmins, Non-Hindu, Illiterate mother, no land ownership, and no motor vehicle ownership.

Table 6: Years of Exposure to Contaminated Water and Educational Outcomes among Non-Repeaters

	(1) Math Scores	(2) Verbal Scores	(3) Absent Days	(4) CGPA
Years Exposed	-0.006	0.003	0.222*	-0.075**
	(0.005)	(0.007)	(0.119)	(0.038)
Observations	2,828	2,126	2,828	2,742
R-squared	0.426	0.337	0.204	0.384

Robust standard errors clustered at the habitation level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All specifications include habitation, grade and school fixed effects. The main explanatory variable is years of exposure to contaminated water as shown in equation 1. The other controls are age, gender, religion, caste, assets, mother's education. The omitted categories of the control dummies are Brahmins, Non-Hindu, Illiterate mother, no land ownership, and no motor vehicle ownership.

Table 7: Years of Exposure to Contaminated Drinking Water and Educational Outcomes: Distance to the Water Treatment Plant (WTP)

Educational outcomes	(1) Early Treated (less than median distance from WTP)	(2) N	(3) Late Treated (greater than median distance from WTP)	(4) N
Absenteeism	0.234	1506	0.287*	1834
	(0.177)		(0.151)	
Grade Repeat	0.037**	1506	0.037***	1834
	(0.016)		(0.011)	
CGPA	-0.134**	1454	-0.110***	1777
	(0.058)		(0.037)	
Math Scores	-0.011	1505	-0.011	1833
	(0.007)		(0.007)	
Verbal Scores	-0.002	1069	0.001	1377
	(0.012)		(0.008)	

Robust standard errors clustered at the habitation level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All specifications include habitation, grade and school fixed effects. The main explanatory variable is years of exposure to contaminated water. Columns 1 and 3 are divided by the median distance of the habitation from the Water Treatment plant (WTP) of respective Zones (Zone I & Zone II). Distance is measured in kms. The other controls are age, gender, religion, caste, assets, mother's education. The omitted categories of the control dummies are Brahmins, Non-Hindu, Illiterate mother, no land ownership, and no motor vehicle ownership.

Table 8: IHDS Descriptive Statistics and District Level Control Variables

	Mean	Std. Dev.
No. of days absent in last month	4.5	5.63
Arsenic (micrograms/L)	105.91	276.88
Percentage clayey soil	28.68	9.36
Hours spent per week doing homework and private tuition	7.88	6.35
Family background characteristics:		
Parental education in years	5.50	4.84
Household monthly consumption expenditure (In Rupees)	16,047.12	7,416.761
Hindu	0.75	0.43
Scheduled Caste/Scheduled Tribe	0.43	0.49
Other Backward Class	0.32	0.47
Individual Characteristics:		
Age	11.79	3.52
% Male	0.53	0.49
District level control variables:		
Female per 1000 males	930.38	44.52
Iron (mg/litre)	1.58	2.26
Ratio rice to wheat (Million tonnes)	1,319.78	6,258.17
Urbanization Rate	27.11	8.70
Literacy Rate	71	8.40
District level Education controls:		
Gross enrolment (government)	341,561	272762
Gross enrolment (private)	129,249	132,913
Number of schools (government)	2,244	1,291
Number of schools (private)	565	545

The sample size is 14,346. The table provides summary statistics of variables specified in Equation (3) and (4).

Table 9: First Stage Regression in IDHS

	Arsenic (microgram/litre)
% Clayey soil	0.009***
	(0.000)
State F.E.	Yes
F-statistic	37.73
Observations	14,346

Robust standard errors (SE) clustered by PSU (village/neighborhood/town level). Significant at ***1%,**5%,*10%. Independent variable is percentage of clayey soil in a district as shown in equation 4. Arsenic is measured in milligrams per litre. Regression includes state fixed effects and district level controls for sex ratio, pattern of cultivation, iron, urbanisation, gross enrolment, number of schools and literacy. Other individual and family related controls are age, gender, religion, caste, parental education and household consumption expenditure.

Table 10: OLS and IV Estimates of Arsenic on Monthly School Absenteeism

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV
	Full sample	Full sample	Full Sample	Full Sample	6-13 year olds
Arsenic	0.044	3.142**	3.959**	2.410*	2.719*
	(0.493)	(1.26)	(1.51)	(1.41)	(2.102)
Male				-0.264	-0.256
				(0.169)	(0.218)
Male*Arsenic				3.238**	3.120
				(1.459)	(1.966)
Individual characteristics	Yes	Yes	Yes	Yes	Yes
Parental characteristics	Yes	Yes	Yes	Yes	Yes
Education factors	Yes	No	Yes	Yes	Yes
Geographical factors	Yes	No	Yes	Yes	Yes
Economic factors	Yes	No	Yes	Yes	Yes
Observations	14,346	14,748	14,346	14,346	9,481

SE clustered by PSU (village/neighborhood/town level). *** Significant at 1%, ** significant at 5%, * significant at 10%. Arsenic is measured in milligrams per litre. Absenteeism is measured in number of days in a month. OLS regression corresponds to Equation (3). Regression includes state fixed effects and district level controls for sex ratio, pattern of cultivation, iron, urbanisation, gross enrolment, number of schools and literacy. Other individual and family related controls are age, gender, religion, caste, parental education and household consumption expenditure.

Table 11: IV Estimates of Arsenic on Health Outcomes

	(1) Diarrhoea in the last 30 days	(2) Diarrhoea in the last 30 days	(3) Diabetes	(4) Respiratory illness in last 30 days	(5) Any Treatment taken in last 12 months
	Entire sample (6 to 19)	Older Children (14 to 19)	Adults (Above 19)	Adults (Above 19)	Adults (Above 19)
Arsenic	0.030	0.085**	0.026**	0.049*	0.065**
	(0.022)	(0.040)	(0.012)	(0.027)	(0.030)
Male	0.001	-0.001	0.002	-0.036***	-0.064***
	(0.004)	(0.006)	(0.002)	(0.008)	(0.008)
Male*Arsenic	-0.002	-0.030			
	(0.031)	(0.047)			
Individual background	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes
Geographical factors	Yes	Yes	Yes	Yes	Yes
Economic factors	Yes	Yes	Yes	Yes	Yes
Observations	14,346	4,865	11,255	11,255	11,255

SE clustered by PSU (village/neighborhood/town level). *** Significant at 1%, ** significant at 5%, * significant at 10%. Arsenic is measured in milligrams per litre. Table presents IV estimates where all outcome variables are binary. Regression includes state fixed effects and district level controls for sex ratio, pattern of cultivation, iron, urbanisation and literacy. Other individual and family related controls are age, age-squared, religion, caste, parental education and household consumption expenditure.

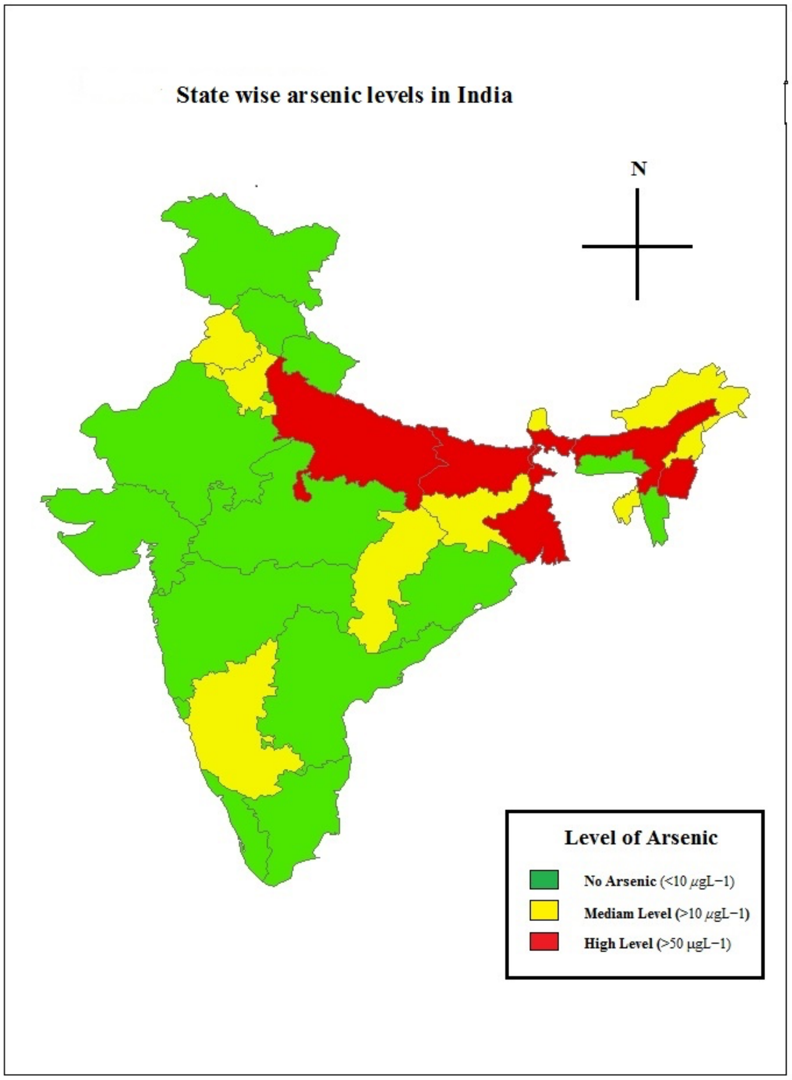
Table 12: Reduced Form Estimates between Clayey Soil and School Absenteeism Across Arsenic and Non-Arsenic Districts

	(1) Full Sample	(1) Arsenic Districts	(2) Non Arsenic Districts
Clayey Soil	0.038**	0.061***	0.027
	(0.012)	(0.018)	(0.019)
State Fixed Effects	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
District Level Controls	Yes	Yes	Yes
Observations	14,346	4,454	9,892

SE clustered by PSU (village/neighborhood/town level). *** Significant at 1%, ** significant at 5%, * significant at 10%. Absenteeism is measured in number of days in a month. Regression includes state fixed effects and district level controls for sex ratio, pattern of cultivation, iron, urbanisation, gross enrolment, number of schools and literacy. Other individual and family related controls are age, gender, religion, caste, parental education and household consumption expenditure.

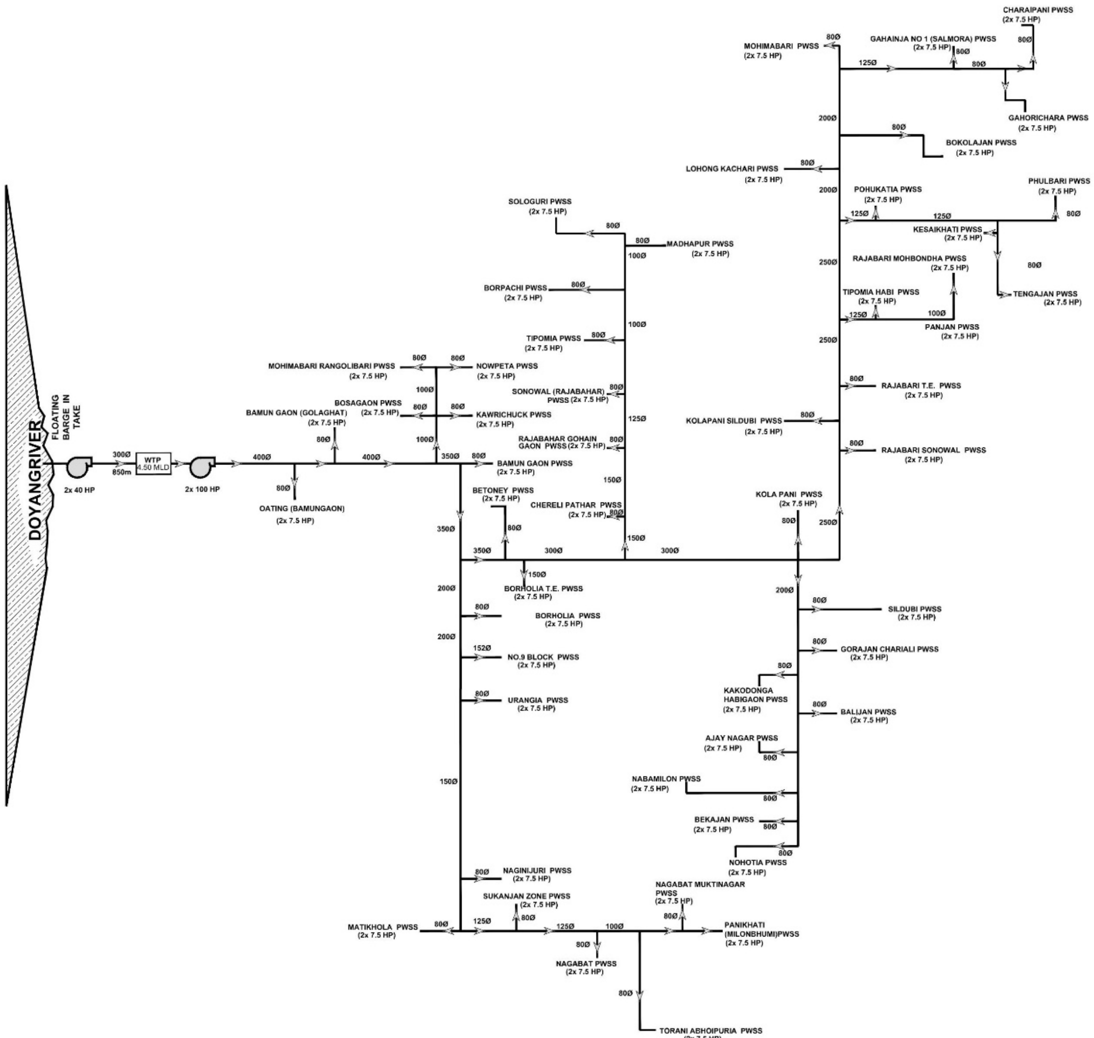
Appendix

Figure A.1: Geographical Distribution of Arsenic Levels Across States of India



Source: Authors calculation using Central Ground Water Board report data (2016).

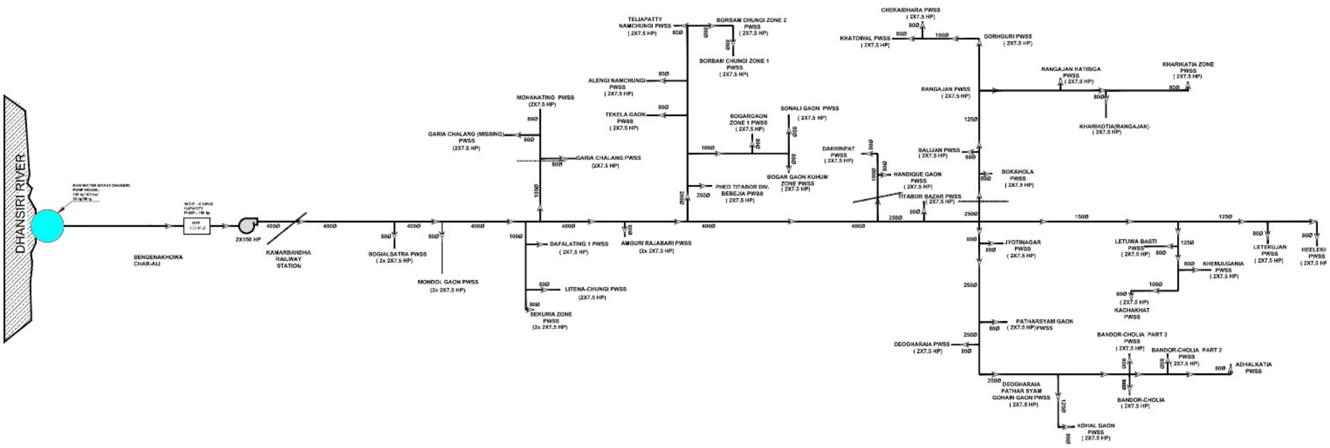
Figure A.2: Location of Doyang River, Water Treatment Plant (WTP) and Pipe Water Supply Schemes (PWSS) to villages in Titabor



Notes: Administrative map of WTP and water supply pipelines connected to the Doyang River (Source: PHED Titabor).

HP-Horse Power; MLD- Million Liters per Day; ∅ Diameter of the pump

Figure A.3: Location of Dhansiri River, Water Treatment Plant (WTP) and Pipe Water Supply Schemes (PWSS) to villages in Titabor



Notes: Administrative map of WTP and water supply pipelines connected to the Doyang River (Source: PHED Titabor).
HP-Horse Power; MLD- Million Liters per Day; Ø Diameter of the pump

A.0.1 Description of Data and Variables (IHDS)

Name of Variable	Data Source	Years	Descriptives
School absenteeism	IHDS-II	2011-12	Table 5
Household consumption expenditure	IHDS-II	2011-12	Table 5
Religion	IHDS-II	2011-12	Table 5
Caste	IHDS-II	2011-12	Table 5
Age	IHDS-II	2011-12	Table 5
Gender	IHDS-II	2011-12	Table 5
Parental education in years	IHDS-II	2011-12	Table 5
Hours spent per week doing homework/private tuition	IHDS-II	2011-12	Table 5
Arsenic	CGWB	2011-12	Table 5
Clayey soil	HWDS	2011-12	Table 5
Sex ratio	Census of India	2011	Table 5
Iron	CGWB	2011	Table 5
Rice/wheat production	MAFW	2011	Table 5
Urbanisation	Census of India	2011	Table 5
Literacy	Census of India	2011	Table 5
Gross enrolment (government)	DISE	2011	Table 5
Gross enrolment (private)	DISE	2011	Table 5
No. of schools (government)	DISE	2011	Table 5
No. of school (private)	DISE	2011	Table 5
Nitrate	CGWB	2011	Table A.0.3
Lead	CGWB	2011	Table A.0.3
Fluoride	CGWB	2011	Table A.0.3
Rainfall	IMD	2011	Table A.0.3
Minimum temperature	IMD	2011	Table A.0.3
Maximum temperature	IMD	2011	Table A.0.3
Potassium	MAFW	2011	Table A.0.3
Nitrogen	MAFW	2011	Table A.0.3
Phosphorous	MAFW	2011	Table A.0.3
Male labour participation (agriculture)	Census of India	2011	Table A.0.3
Female labour participation (agriculture)	Census of India	2011	Table A.0.3

IHDS: India Human Development Survey; CGWB: Central Ground Water Board; HWDS: Harmonized World Database Software; MAFW: Ministry of Agriculture and Farmer's Welfare; DISE: District Information System for Education; IMD: Indian Meteorological Department.

A.0.2 Self-Collected Survey Data Description (Titabor)

Name of Variable	Data Source	Years	Data
Respondent characteristics			
School absenteeism	School level survey	2018	Table 1
Percentage scored in previous grade	School level survey	2018	Table 1
Numeracy scores	School level survey	2018	Table 1
Verbal scores	School level survey	2018	Table 1
Major source of drinking water	School level survey	2018	Table 1
Age	School level survey	2018	Table 2
Male	School level survey	2018	Table 2
Religion	School level survey	2018	Table 2
Caste	School level survey	2018	Table 2
Structure of the dwelling	School level survey	2018	Table 2
Assets ownership	School level survey	2018	Table 2
Mother's education	School level survey	2018	Table 2
Awareness of contaminants	School level survey	2018	Table 2
School level characteristics			
No. of teachers	School level survey	2018	Table A.0.4
Teaching experience	School level survey	2018	Table A.0.4
Class size	School level survey	2018	Table A.0.4
Playground availability	School level survey	2018	Table A.0.4
Hours of electricity supply	School level survey	2018	Table A.0.4
Availability of toilet facility	School level survey	2018	Table A.0.4
Availability of library facility	School level survey	2018	Table A.0.4
Availability of playground facility	School level survey	2018	Table A.0.4
Village level characteristics			
Population	Census of India	2011	Table A.0.5
Children's population (0-6 years)	Census of India	2011	Table A.0.5
Schedule caste (percentage)	Census of India	2011	Table A.0.5
Schedule tribe (percentage)	Census of India	2011	Table A.0.5
Literacy	Census of India	2011	Table A.0.5
Total workers	Census of India	2011	Table A.0.5
Marginal workers	Census of India	2011	Table A.0.5
Population	PHED*	2011	Table A.0.5
Safe water scheme rollout	PHED	2009-2018	Table A.0.5

*Public Health Engineering Department, Ministry of Drinking water and Sanitation.

Table A.0.3: District level Variables (IHDS)

Variable	Mean	Std. Dev.
Iron (mg/litre)	1.63	2.42
Nitrate (mg/litre)	60.66	90.31
Fluoride (mg/litre)	0.52	1.02
Lead (mg/l)	1.95	6.53
Nitrogen (Kilogram/hectare)	27633.40	26607.39
Phosphorus (Kilogram/hectare)	12472.62	13998.89
Potassium (Kilogram/hectare)	4261.75	6203.05
Rainfall (millimetres)	80.39	60.12
Maximum temperature (degree Celsius)	9.43	16.82
Minimum temperature (degree Celsius)	1.74	4.56
Pattern of cultivation (ratio of rice/wheat in mill/tn)	1777.28	15899.70
Literacy rate (percentage)	68.98	8.48
Sex ratio (per 1000 males)	936.28	44.90
Annual Consumption expenditure (INR)	169859.60	63435.90
% Male labour force participation in agriculture	19.52	17.44
% Female labour force participation in agriculture	6.68	9.91
N=205		

Table A.0.4 Descriptive statistics on Aggregate School Quality Measures

Variable	Mean	Std. Dev.
Experience of teachers (years)	15.49	13.28
No. of school teachers	6.68	7.07
Class size	24.23	21.46
Playground (proportion)	0.49	0.50
Duration of power cut (hours)	2.816	1.592
Semi-flush/flush toilets (proportion)	0.82	0.39
Accessibility to library (proportion)	0.58	0.49
Proportion of Scheduled Caste/Scheduled Tribe/Other Backwards Class	0.82	0.39
N=117		

Accessibility to playground, semi-flush/flush toilets and library are binary variables. Schools with less than 60 percent of Backward/marginalised (OBC/SC/ST) are coded as 0 and is the omitted category.

Table A.0.5 Village Characteristics for Titabor Block

Variable	Mean	Std. Dev.
Population	1111.4	686.7
No. of children (0-6 years)	126.9	94.5
Scheduled caste	66.6	144.4
Scheduled tribe	152.9	205.5
Literacy rate	80.6	12.7
Total workers	544.8	360.5
marginal workers	208.3	192.2
No. of years of piped water scheme	3.63	2.63
Total households	237.82	145.76

N=158. According to census of India (2011), total workers are defined as sum of marginal workers and main workers. Main workers is defined as a person who has worked for major parts of the reference period (i.e. in the last one year preceding the date of enumeration) in any economically productive activity. Marginal workers is defined as a person who worked for less than six months of the reference period (i.e. in the last one year preceding the date of enumeration) in any economic activity.

Table A1: Clayey Soil and District Level Characteristics

	Clayey soil
Iron (mg/litre)	-0.004
	(0.025)
Fluoride (mg/litre)	0.011
	(0.132)
Nitrate (mg/litre)	0.111
	(0.881)
Lead	0.051
	(0.085)
Nitrogen (Kilogram/hectare)	319.1
	(283.3)
Phosphorus (Kilogram/hectare)	331.8
	(167.6)
Potassium (Kilogram/hectare)	-51.38
	(57.64)
Weather:	
Rainfall (millimetres)	-0.803
	(0.555)
Maximum temperature (degree Celsius)	-0.352
	(0.212)
Minimum temperature (degree Celsius)	-0.030
	(0.065)
Demographic and Economic Factors:	
Ratio of rice to wheat (in millions of tonnes)	-50.92
	(87.19)
Literacy	0.028
	(0.095)
Sex ratio (per 1,000 females)	-0.151
	(0.378)
Per capita expenditure (INR)	750.8
	(808.5)
Male to Female Ratio labor force participation (agriculture)	0.105
	(0.109)
	(0.044)
State fixed effects	Yes

Table A.1 reports the coefficient on clayey soil, from the regression of reported district level variables on the % of clayey soils in a district and state fixed effects. Standard errors are clustered at the district level. N=149 districts

Table A2: Correlation Between Year of Water Supply and Aggregate School Quality Measures

Aggregate School Quality Measures	Years of safe water supply in habitation
Teaching experience (years)	0.000
	(0.014)
No. of teachers	0.025
	(0.031)
Class size	-0.005
	(0.010)
Playground availability	-0.007**
	(0.003)
Electricity access (hours)	0.063
	(0.091)
Toilet	-0.011
	(0.263)
Library	0.002
	(0.004)
Proportion of Scheduled Caste	0.351
	(0.329)
Population	0.001*
	(0.000)

N=79. Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable is an index for the year in which a habitation got access to safe water ranging from 1 (for 2008) to 9 (for 2018) years.

Table A3: Correlation Between Awareness and Aggregate School Quality Measures

Aggregate School Quality Measures	Awareness of Contaminants
Teaching experience (years)	0.023
	(0.111)
No. of teachers	-0.242
	(0.242)
Class size	0.190***
	(0.074)
Playground availability	0.009
	(0.026)
Electricity access (hours)	-0.964
	(0.707)
Toilet	-0.040
	(2.034)
Library	-0.464
	(0.029)
Proportion of Scheduled Caste	1.221
	(2.545)
Population	0.004
	(0.003)

N=79. Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable measures the proportion of respondents with awareness regarding presence of contaminants. Contaminants include arsenic, nitrate, fluoride, and iron.

Table A4: Years of Exposure to Contaminated Drinking Water and Health Outcomes

	(1)	(2)	(3)	(4)	(5)
	Self-Skin diseases	Self-Other diseases	Parents Skin diseases	Parents other diseases	Parents long term diseases
Years Exposed	0.003	0.006	-0.006	0.004	0.009
	(0.009)	(0.008)	(0.008)	(0.010)	(0.011)
Observations	3,340	3,340	3,340	3,340	3,309

Robust standard errors clustered at the habitation level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All specifications include habitation, grade and school fixed effects. The table shows the effect of years of exposure to contaminated water on health outcomes (corresponding to Equation (1) where the left hand side variable is health outcomes). The other controls are age, gender, religion, caste, assets, mother's education. The omitted categories of the control dummies are Brahmins, Non-Hindu, Illiterate mother, no land ownership, and no motor vehicle ownership. Self-Skin diseases is binary variable which takes value 1 if the respondent is suffering from any skin diseases and 0 otherwise. Self-other diseases is binary variable which takes value 1 if the respondent is suffering from fatigue/unhealthy weight loss/joint pain/ nausea, vomiting/ stomach problems and 0 otherwise. Parent's Skin diseases is binary variable which takes value 1 if the respondent's mother or father is suffering from skin diseases and 0 otherwise. Parent's other diseases is binary variable which takes value 1 if the respondent's mother or father is suffering from fatigue/unhealthy weight loss/joint pain/ nausea, vomiting/ stomach problems and 0 otherwise. Parent's long term diseases is binary variable which takes value 1 if the respondent's mother or father is suffering from chronic or fatal diseases such as organ damage, cancer, mental illness, muscular/bone damage or respiratory diseases and 0 otherwise.

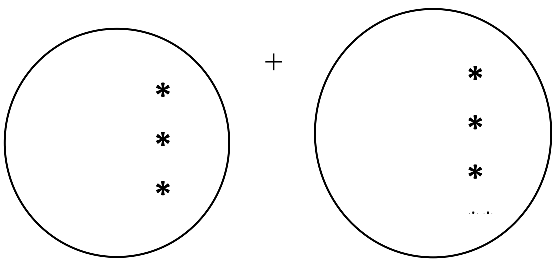
A5 Titabor Survey Cognitive Skill Questions

Mathematical Ability of third grade students

Q1. Addition: 45+38=?
1) 73 2) 38 3) 90 4) 83

Q2. Multiplication: 30*2=?
1) 60 2)50 3)55 4)70

Q3. Counting



= ?

1) 5 2) 8 3) 6 4) 7

Verbal ability of fifth grade students

Read the passage below and answer the questions that follow:

She had some sweets that she wouldn’t share,
She had a book that she wouldn’t lend,
She wouldn’t let anyone play with her doll,
She is nobody’s friend!

He had some toffee and ate every bit
He had a tricycle he wouldn’t lend,
He never let anyone play with his train,
He is nobody’s friend!

But I will share all of my sweets with you,
My ball and my books and my games I will lend,
Here’s half my apple and half my cake
- I am your friend!
ENID BLYTON

Q1. What are the things that the girl does not want to share?
01) Book, doll and sweets 02) Balloons
03) Bicycle 04) None of the above

Q2. What are the things that the boy does not want to share?
01) Books 02) Tricycle, toffee and train
03) Balloons 04) None of the above

Q3. What does the child in the last paragraph want to share?
01) Sweets, apple and cake 02) Ball and Books
02) 03) Games 04) All of the above

Mathematical Ability of fifth grade students

Q1. Addition: 7010 + 2699=?
01) 9799 02)9699 03)9709 04)9609

Q2. What is the difference between 500.2 and 499.101?
01) 1.099 02)1.990 03)1.109 04)1.101

Q3. The digits 3 and 4 of the number 354 are inter-changed to form a new number. What is the difference between the new number and the original number?
01) 199 02)101 03)109 04)99

Verbal ability of 8th grade students

Analyse the picture given below and answer the following questions below:

Q1. The advertisement is about the importance of which of the following?
01) Benefits of nutritious food 02) Benefits of yellow peas Dal
03) Reasonably priced food 04) All of the above

Q2. Who has issued this advertisement?
01) Kendriya Bhandar
02) Ministry of Consumer Affairs, Food and Public Distribution
03) Central Food Technology Research Institute
04) Mother Dairy.

Q3. What is the meaning of the word "Nutrition"?
01) Process of purchase of costly food
02) Food rich in unhealthy fats
03) Process of providing or obtaining food for health and growth
04) Process of making handmade food

<u>Mathematical Ability of eighth grade students</u>

Q1. Three exterior angles of a quadrilateral are 70 degree, 80 degree and 100 degree. The fourth exterior angle is:
1) 70 degree 2) 80 degree 3) 100 degree 4) 110 degree

Q2. If $(x + 8) = 15$ then the value of x is:
1)10 2) 11 3) 7 4) 15

Q3. If $(2x + 1)/(x + 3)$ =1 then the value of x is:
1) 2 2) 3/2 3) 1 4) -1