The Effects of Skill-Based Health Education
--- A Randomised-Controlled Intervention in Primary Schools in Rural Bangladesh

Makiko Omura

Faculty of Economics
Meiji Gakuin University

Discussion Paper No.24-01
April 2024

The Effects of Skill-Based Health Education --- A Randomised-Controlled Intervention in Primary Schools in Rural Bangladesh

Makiko Omura*

Abstract

This paper investigates the impact of skill-based health education (SBHE) on school hygiene status, and child health and hygiene practices, and health status in rural Bangladesh. A weekly session of SBHE was given to primary schools through a randomised-controlled trial (RCT) by locally recruited trained para-teachers for the duration of one year. We found that SBHE exerted positive effects on healthy practice and behavioural change at both school and child levels, including better school hygiene infrastructure management. Improvements in cold-related symptoms were also observed among the treated children. The analysis suggested a strong general improvement trend in healthy practices regardless of the treatment status. Several measures of inter-school spillovers were employed, which provided robust evidence of beneficial externalities pertaining to healthy practices at both school and child levels. A cost-effectiveness analysis suggested that our SBHE was cost-effective, particularly incorporating spillover effects. A cross-cutting soap provision treatment, although implemented imperfectly, did not suggest any standalone positive health-related effects. This study is registrated in the AEA RCT registry (No.0004265) and the ISRCTN registry (No.18002856).

* Department of Economics, Faculty of Economics, Meiji Gakuin University, 1-2-37 Shirokanedai, Minato-ku, Tokyo 108-8636, Japan. makiko@eco.meijigakuin.ac.jp

This research project is executed in cooperation with Save the Children, Inc. and was enabled by the research grants, the Japanese Grant-in-Aid for Scientific Research (No. 23402033) from the Japan Society for the Promotion of Science and from Nomura Foundation, and by in-kind contribution and collaboration with Save the Children (SC), Inc. (originally SC, USA), Dhaka Office. I am particularly grateful to Mohini Venkatesh, Ikhtiar Uddin Khandaker, and Ataur Rahman, who were at the SC at the time of intervention, Shuaib Muhammad and Ziaul Hasan of SURCH in Dhaka, as well as all the recruited project staffs and survey staff for their dedicated work in enabling this project. My gratitude also goes to the officials of the Bangladesh Ministry of Primary and Mass Education and the Director General of Health Services officials, officials of the District of Jhenaidah, as well as the head teachers and all other people who enabled and supported this project in one way or the other. Gratitude extends to all teachers, children and their families for their valuable contributions participating in this project. I also wish to express my deepest condolences for the premature death of Taslima Khatun Gini, one of our field investigators in Jhenaidah, at the time of the endline survey. I am grateful to my colleagues and various seminar and conference participants for their valuable comments. I am also thankful for the advice provided for field experiment design by Pascaline Dupas, Rachel Glennerster, Marc Shotland, and Iqbal Dhaliwal of Abdul Latif Jameel Poverty Action Lab, MIT. All errors are our own.

Introduction

This research evaluates the effect and cost-effectiveness of skill-based health education (SBHE) in inducing healthier and hygienic practices among schools and pupils, and additionally improved health in pupils through a randomised-controlled trial (RCT) in primary schools in rural Bangladesh. In Bangladesh, despite concerted efforts by both the government and international bodies to enhance health outcomes among school-age children through the provision of anthelminthic drugs, nutritional supplements, and improved sanitary infrastructure such as latrines, significant health challenges persist. According to the Information provided by the Bangladesh Directorate General of Health Services at the time of study [1], the major illnesses affecting school-age children were diarrhoea, micronutrient deficiencies, malnutrition, intestinal helminths, pneumonia, acute respiratory infection, typhoid and scabies. Since then, school-based deworming programme as well as vaccination programme progressed. Nonetheless, illnesses such as diarrhoea, malnutrition, pneumonia and acute respiratory infection are still of concern [2]. Likewise, while current data shows some improvements in sanitation and hygiene, only 56% of schools have access to basic hand hygiene facilities, which can exacerbate health risks [3]. This highlights the ongoing need for robust public health strategies and interventions to secure the well-being of Bangladesh's children.

While the major interventions for child health improvements in such context is goods provision, our primary purpose is to examine the effectiveness of school-based SBHE focusing on skill acquisition and habit development by integrating skill-building sessions in teaching, because the effects of acquired skill and habit can be sustained without perpetual external inputs. The SBHE aims to improve the health-related school environment, and children's health and health-related knowledge, attitudes, practices, and behaviour (KAPB). Healthy school environment and habit imply that school hygiene status is better maintained. One of the major reasons for the lack of hygiene infrastructure in developing countries beyond limited budget is its inadequate management and resulting short lifetime after its provision. The current situation in Bangladeshi schools suggests that the already poorly

available infrastructural resources are poorly managed and not fully utilised – newly built latrines become unusable and discarded after several years. Latrines are often locked and/or in bad conditions, and soaps are unavailable or available only at the teachers' room upon request. A detailed case study in primary schools in Niger by Tomokawa, Kasai, and Kobayashi (2005) also highlights this problem. Changing default behaviour towards healthy and hygienic one can not only contribute to better infrastructure maintenance but would be much less costly than building a new infrastructure, and realisable in a shorter time with possibly life-long effects. A cluster-RCT focused on enhancing the upkeep of shared toilets in Dhaka's urban slums demonstrated that promoting behaviour change along with the provision of low-cost simple goods resulted in improved maintenance of those hygiene facilities [5]. Healthy habits can reinforce any other positive impacts of supply-based interventions, even inducing demand and habits for infrastructure maintenance activities.

Although the importance of school health education is generally recognised [6], relatively few rigorous, comprehensive evaluation has been conducted possibly due to difficulties in measuring its impact. Also, relatively less attention has been paid to health of primary school-aged children compared to pre-school children or adolescents [7]. Among the few evaluation studies, health education was often found to be ineffective in comparison to provision of drugs and/or supplements which was relatively cheap with its benefit materialising in a shorter time-period [8–10]. Nonetheless, there have been some evidence of positive health education focusing on specific theme, such as oral hygiene education [11,12] and handwashing [13,14]. The SBHE differs from typical lecture-based health education as it focuses on the development of context-relevant knowledge, attitudes and skills in children using a variety of participatory learning methods. An SBHE study conducted in Tanzania showed significant behavioural impact such as reduced sugary food consumption, more frequent teeth-brushing and improved brushing-skills compared to those who received conventional oral health education [15]. A randomised pairwise evaluation of SBHE regarding schistosomiasis in Egyptian primary schools found significant improvements not only in knowledge and attitudes but also in

schistosomiasis infection [16]. Although their findings cannot be taken at face value because of the study's less rigorous experimental design, the SBHE approach is suggestive. A clustered-RCT intervention of weekly handwashing promotion targeting households in a Karachi squatter settlement in Pakistan, together with soap provision, resulted in major reductions in the incidence of pneumonia, diarrhoea, and impetigo [13]. Another clustered-RCT of respiratory hygiene education and hand sanitizer supply showed successful reduction in influenza among schoolchildren in urban Bangladesh [17]. A study comparing environmental nudging and intensive hygiene education in 20 Bangladeshi primary schools showed that both methods improved child handwashing behaviour, although the provision of hand-washing stations alone had no positive effect [14]. Education projects on HIV/AIDS prevention in both developed and developing countries offered evidence of official educational curriculum to be ineffective in contrast of simple but relevant information provision being effective [18,19]. They suggest that context-relevant information and practical skills can be the key elements of successful SBHE, for which school can be an effective setting as it allows targeting many pupils at the same time inducing economies of scale and peer effects.

While successful health education intervention cited above often provided goods and/or infrastructural supply along with education, our study employing cross-cutting intervention of SBHE and soap provision intends to assess the impact of health education, with/without goods provision. Healthy skills are introduced and practiced as a part of school curriculum. The reason for intervening with school children is not only to give attention to those age group whose heath is paid relatively less attention but also in expectation that they are more amenable to changes in habits. Children are expected to form healthy habits through regular SBHE sessions at school with visual aid and repeated learning-by-doing together with classmates, receiving context-relevant advice. The importance of repeated learning is attested by [12]. While a child as an individual may feel it cumbersome in the beginning to wash hands, wear shoes, dispose waste in a waste box or clean their classroom, etc., they are induced to repeat these actions which can become habitual. Beneficial peer-effects are also

expected among the schoolmates as has been suggested by different works [20–24], although evidence for negative peer-effects also exist [25,26]. Despite habit and behavioural inertia resistant to change, psychological research has shown that habit formation could be achieved in two to three months through simple and sustainable behavioural advice [27]. While habituation of elaborate actions such as physical exercise is believed to require overcoming time inconsistency, where one's short-run actions are perceived as suboptimal from one's long-run perspective [28,29], simpler action such as washing hands or putting on a seatbelt can be automated with a lesser degree of time inconsistency. This is because the action's immediate cost is relatively low, and the realisation of its benefit, at least in part, is also immediate. A short time lag can also contribute to action exertion since the net present value of the future benefit is not heavily discounted relative to the present cost. Thus, habituation or automaticity is faster and stronger for simple actions [27,30,31]. Through habit, the behaviour becomes 'second nature,' which is cognitively efficient and does not require conscious effort. Thus, making healthy behaviour habitual is the crucial aim of SBHE.

In addition to assessing the effect of SBHE, we conduct a cost-effectiveness analysis. The authors are aware of limited study, such as one by [9], which explicitly evaluates the cost-effectiveness of health education, as most studies do not offer cost analysis. In this regard, our claim that school-based SBHE and the use of para-teacher to be cost-effective vis-à-vis regular-type of health education is examined to a limited extent. The rest of the paper describes the project and intervention design, data, analysis with empirical model and estimation results with additional spillover considerations and cost-effectiveness analysis. In the conclusion, we summarise and succinctly discuss the findings from the study.

Methods

Project and intervention design

The project was conducted in Jhenaidah District (*zila*), which was a target district of Save the Children (SC)'s PROTEEVA (Promoting Talent Through Early Education) project. However, our intervention specifically took place in two non-PROTEEVA sub-districts (*upazila*), Moheshpur and Kodchandpur, spanning areas of 21.16km^2 and 20.16km^2, respectively. These neighbouring sub-districts were chosen due to their need logistical convenience, given the existing SC offices and local NGO partnerships in the region. To prevent any crossover effects, our project did not involve any PROTEEVA-targeted schools/communities. The intervention was executed for over 12 months from 1 March 2012 to 25 March 2013, preceded by a four-month period of baseline data collection for which recruitment commenced on 13 October 2011 and concluded on 29 November 2011. The endline data collection was conducted for five months after the project's completion, from 7 April 2013 to 6 July 2013, complemented by further follow-up and data verification processes.

The proposed project applied a treatment-control pre-post evaluation based on a cross-cutting randomisation design of SBHE (HE) and a soap-provision (SP) intervention. The unit of intervention was school, and 180 randomly chosen schools out of total of 204 primary schools were stratified according to the school type—government primary school (GPS) and registered non-government primary school (RNGPS). A cross-cutting HE- SP treatments were then randomly assigned to 180 schools stratified by two school-type stratified. Thus, four groups (HE, SP, HESP, and control) with 45 school each were randomly chosen. In Bangladesh, the main types of schools GPS and RNGPS which, despite the name, are actually government-funded, although they receive only a fraction of government funds received by GPS. Thus, RNGPS are disadvantaged in terms of resources and school infrastructure. Therefore, it is important to stratify by school type. Half of the HE-treatment and control schools was assigned to an additional SP-treatment for one year, where six soap bars were monthly given to SP-treatment schools and three soap bars to randomly chosen children within the schools. The

SP-treatment was implemented to assess whether goods provision per se or interactively with HE could induce healthy behaviour, in this case, handwashing. For SP-treatment, no health advice was provided but schools and pupils were assigned to receive six and three small soap bars monthly, respectively. The SP-treatment had another role in mitigating possible Hawthorne effects, especially given our primary objective of discerning the impact of SBHE. Hawthorne and John Henry effects are changes in the behaviour of HE-treatment and control groups, respectively, due to the sheer fact of evaluation taking place. In order to mitigate John Henry effects among the pure controls, the project provided game boards to the control schools, which did not affect the measurability of our intervention. Another way to amend these effects would be to collect data for some time, especially after the completion of formal evaluation, but that was outside the scope of this project.

The project was approved and supported by the Bangladesh Ministry of Education, Directorate of Primary and Mass Education in Dhaka, the Directorate of General of Health Services, and the District Primary Education Officer in Jhenaidah, *Upazila* Education Officer the schools in the Moheshpur and Kodchandpur *upazilas*. The project was designed to provide schools and school children with needed health education and no particular health or physical risk was envisaged. This study obtained an ethical approval by the Research Integrity Review Board of Meiji Gakuin University. The field experiment was registered with the American Economic Association's registry for randomised controlled trials (AEARCTR-0004265) and the ISRCTN registry (ISRCTN-18002856). The registrations were done retrospectively given the fact that AEARCTR was not yet established at the time of initial participant recruitment, and that health education intervention was not considered to be a clinical trial by the authors at the time of initial participant recruitment concerning the ISRCTN registry. We adhered to Consolidated Standards of Reporting Trials guidelines for reporting results of cluster randomised controlled trials (see Supporting Information S1 File for CONSORT checklist [32]).

Skill-Based Health Education Strategies

SBHE projects are often considered expensive, labour intensive, and requiring long-term involvement, which many practitioners find unsustainable and difficult to replicate to scale, while their impact is hard to discern. The project was thus designed with the following key features: (1) weekly health education sessions using participatory learning methods to promote habit-formation; (2) use of para-teachers instead of primary school teachers as the SBHE facilitators, given their cost-effectiveness and the incentive problems expected among the incumbent teachers; and (3) the use of a mobile projector to assist pedagogy and skills learning and to motivate both pupils and teachers. The device could be easily charged in advance when there was power supply, which was recommended due to Bangladesh's unreliable power supply.

The health education session consisted of 26 modules as shown in Table 1. The educational contents utilised many images and videos relevant to the Bangladesh context, which were made appropriate for primary school pupils, as well as learning-by-doing sessions, for handwashing, brushing teeth, saline making, latrine usage, and latrine cleaning, aiding children to acquire health-related KAPB. Project staff also recorded various videos, including one featuring a smartly dressed male staff member demonstrating the correct method of cleaning a latrine. The use of these materials aimed to provide practical guidance and to help diminish or eliminate biases against 'dirty jobs.' Each week, a one-hour education session was conducted during the time slot allocated for Physical Education (PE) for each grade. Typically, PE periods were used as free time for children to play. If playing around is assumed to have positive health effects, which may well be the case, the estimated HE effects on health may be biased toward zero.

Table1. 26 Skill-based Health Education Modules

SBHE modules (26)	Contents
1. personal hygiene (3)	handwashing, hair and nail trimming, bathing, cloth-cleanliness, eye care; teeth-brushing; tooth decay

2. sanitation (3)	keeping premises clean and waste management; latrine cleaning; sanitary latrine use
3. safe water (3)	safe and unsafe water; water purification and preservation; arsenic problem and prevention
4. common illness (6)	fever, cold and cough, diarrhoea cause, diarrhoea prevention and treatment, causes and demerits of worms, deworming
5. nutrition (5)	balanced diet, symptoms of malnutrition, vitamin A, iodine, iron
6. first aid (3)	burns, cuts, fractures
7. injury prevention (3)	burns, drowning, road accidents

Para-teachers conducted five classes daily at a single school and rotated among five different schools throughout the week, resulting in a total of 25 hours of SBHE sessions each week. They were all recruited locally who were either recent college graduate or previous NGO worker looking for jobs. Some of them also had previous experiences in the health/education field. There were several reasons for recruiting them. Firstly, incumbent teachers were already overwhelmed with their existing curriculum, and similar SC projects had faced significant moral hazards when employing these teachers, despite offering extra payment. Meanwhile, the para-teacher strategy proved to be successful in another SC project in Meherpur district of Bangladesh that some schools that had not been supported by SC started to recruit para-teachers with their own initiatives and funds. Also, an RCT study promoting better school climate/environment showed that a new, low-cost lay counsellor produced substantially beneficial effects while a regular teacher produced no effect [33]. Secondly, younger job seekers would be more flexible and eager to learn and teach the SBHE contents that required the application of skill-based teaching methods and of a mobile projector. Indeed, para-teachers considered that going through the SBHE training and practice improved their capacity and skills as prospective teachers. Thirdly, it was considered sustainable cost-wise even after the project, as the cost per para-teacher was 5,000BT per month, approximately 67USD, based on the actual market exchange rate at that time, at 1USD = 74.66BDT, which was marginally lower than the minimum salary of 5,900BT for a regular GPS assistant teacher, yet manageable for the hiring schools, costing them only 1,000BT or 13.4USD per school. Their honorarium was also equivalent to what was being paid by the Meherpur

schools stated above, thus considered sustainable even after the project completion. The para-teachers received 40+ hours of intensive participatory-training in SBHE by BRAC health education specialists which enabled them to deliver specialised contents. A refresher workshop was conducted in the midway for their experience-sharing, feed-back and suggestions.

To date, no comprehensive school health education is offered for primary schools in Bangladesh. The latest health policy states the objective to train at least one teacher per school on health issues [2], however, this kind of arrangement would unlikely be successful for the reasons stated above.

Samples Selection

In the target area, there were 103 GPS and 101 RNGPS at the time of study in 2011-2013. One-hundred eighty primary schools, stratified by school types, were randomly selected from the list of all GPS and RNGPS schools in Moheshpur and Kodchandpur in Jhenaidah District using Excel random classification formula. The process was repeated until statistical nondifference of baseline school characteristics between the groups was ensured. While the intervention was conducted at school level, data were collected at both school and child level. Children surveyed were chosen randomly using seat placement based on the pre-determined randomly selected seat numbers prescribed by the lead author who was also the principal investigator. Surveyors were masked about the treatment status in both baseline and endline surveys. The treatment assignment was done after the baseline data collection, thus all participants were masked about the treatment at the time of baseline survey.

Observations were made for 180 schools at both baseline and endline, and for 7,200 and 9,000 pupils for the baseline and endline surveys, respectively. The sample size calculation at the time of the baseline survey was based on the expected improvement of 0.15 standardised effect size in child health-related indicators, detected with 80% power and 5% significance level, assuming the school-level covariates to explain 2.5% of the variance. The standardised effect size of 0.15 was a conservative estimate yet deemed to be a reasonable degree of improvement. The sample size calculation utilised the intracluster correlation of 0.058, which was derived from SC's baseline survey data in a different

area of Bangladesh in which SC implemented a school health and nutrition project. A sample size of 34 students per schools would have been sufficient. However, to account for partial compliance and/or attrition, which was assumed to be 20%, we collected data for 40 students in grades 1–4 at baseline and 50 students in grades 1-5 per school at endline.

Bangladeshi primary schools are for grades 1–5. Because the sample targets were those who were in primary school during the intervention period, the baseline survey prior to the intervention was conducted for grades 1–4 who would be in grades 2–5 at the time of intervention, and grades 3–6 at the time of the endline survey unless they repeated or dropped out of the class. Additionally, we collected data from ten grade 1or 2 pupils per school in the endline survey, who would have been in grade 1 at the time of the intervention. The endline data were thus collected from grades 1–6, where grade 1 or 2 signified those who had been pre-primary at the baseline, and grade 6 students had already graduated from primary school at the endline. Replacement child data of the same sex and same baseline class were collected for attrition at the endline.

Data collection was done by the Dhaka based survey institute SURCH who received intensive training on the questionnaires, measurement and interview methods, subject random selection method by the lead author and conducted a pilot survey with the authors. For school data, interviews were conducted to headteachers, and observational data were collected with photographs. For child level data, interviews and observational data were collected. All interviews and data collection used structured questionnaires. Prior to the study commencement, participation agreement was obtained from the District Primary Education Officer and school headteachers in the meeting. Parents/ guardians were briefed on the project details and the possibility of their children participating in the survey conducted at the school. They were then asked to provide their consent. At the time of the child survey, written consent was obtained from them using the child assent form. Surveyors were available to assist in completing the forms for the participants when necessary. Children were excluded from the survey if either they or their parents or guardians did not provide consent.

Statistical analysis

We primarily assess the impact of HE intervention on health and health KAPB both at the school and child level. Since the intervention was done at the school level, only intent-to-treat (ITT) effects can be measured at the child level. At each level, the effects of treatment are measured for family-wise summary indices in which the outcome variables y are categorised into families of related variables, as explained below. The family-wise mean-standardised average treatment effects are estimated through seemingly unrelated regressions (SUR) that allow contemporaneous errors to be correlated [34,35]. This analysis accounts for possible cherry-picking caused by the increased likelihood of finding significant results simply due to conducting many regressions [36,37]. We provide additional estimation results for single outcomes in the Supporting Information. Note that the fact that many of the outcome variables y being binary does not pose any particular challenge in obtaining the average causal effect of a random treatment, as the average treatment effect (ATE) exhibits differences in probabilities of $y = 1$ [38]. All statistical analysis was conducted using STATA 16, except for the summary statistics of baseline characteristics which used 'tableone package' in R (ver.4.2.2).

School-level outcomes

We have three school-level outcome families (denoted here with "S"), one primary outcome family and two secondary outcome families. The primary outcome family, (S1) *school hygiene practice & maintenance*, which is considered as the direct indicators of healthy skill-building and healthy behaviour, consists of routine latrine cleaning, latrine cleaning days/week, routine classroom cleaning, classroom cleaning days/week, classroom and corridor cleanness, classroom rubbish bin provision, latrine brush provision, rubbish treatment methods, and soap provision at pupils' handwashing place. While half of the schools were SP-treated, the schools had no instruction nor supervision on how to

use the distributed soap. Two secondary outcome families are indirect outcomes derived from the primary outcomes. One is (S2) *school hygiene infrastructure*, composed of clean accessible sealed-pit latrines for each sex, ratio of clean-usable latrine for each sex, closeness of handwashing facility to the latrines (within 10~11 steps), and easiness of obtaining water by younger children. While sealed-pit latrine can be affected by its structure, particularly water-sealed latrine rather than lid-sealed latrine, it was commonly observed that latrines were not properly water-sealed due to their ill usage and ill maintenance. The number of latrines for each sex included unisex ones. Another family is (S3) children's *schooling* which comprises absence rate, dropout rate, and repetition rate at each grade. A related manuscript by the authors [39] provides a concise analysis at the school level, differing slightly in the outcome-family construct and including additional aggregated child outcomes at the school level.

Base empirical model

The basic construct of the empirical model is the same across school and child-level analyses. While the ITT estimate of average treatment effect (ATE) can be estimated by a simple OLS for a perfect RCT design, an additional vector of control variables **X** for j^{th} of J schools makes the estimates of treatment effects β more precise and accounts for chance differences between groups in the distribution of pre-randomisation traits. In the equation below, Y_j is an outcome and T_j is the treatment status for school j, and D_j is a *school-type* dummy which is used for sample stratification of school types, GPS/RNGPS.

$$(1)\ Y_j = \alpha + \beta \cdot T_j + \varphi \cdot \mathbf{X}_j + \theta D_j + \varepsilon_j.$$

Because of its randomised design, we may assume that random assignment of schools to treatment and comparison groups ensures that the schools in either group are similar in all other respects except for having/not receiving a treatment. To ensure this, school-type stratified randomisation process was repeated until the statistical indifference between treatment and control groups was ensured on the main variables. Nonetheless, there is always a possibility of randomisation being imperfect. Thus, a

difference-in-differences (DID) estimation is conducted to adjust for any pre-existing differences between treatment and comparison schools:

$$(2)\ Y_{jt} = \alpha + \beta \cdot t \cdot T_j + \delta \cdot t + \gamma \cdot T_j + \varphi \cdot \mathbf{X}_{jt} + \theta D_j + \varepsilon_{jt},$$

where t is the time-period, T_j is the treatment status for school j. School-level DID can also be estimated with school fixed effects that control for time-constant unobservable school characteristics.

Child-level outcomes

The primary child-level outcomes concern health-related KAPB change, measured through pre-post questionnaires, measurements, and observations. Health-related KAPB indicators are: (P1) *handwashing practice*: handwashing habit index (frequency and used materials before eating, after defecation, and after playing), frequency of handwashing with soap on each occasion, washing under running water, and correct handwashing procedure; (P2) *dental-care practice*: dentalcare index (frequency); tooth-brushing frequency using brush/branch; combination of used materials such as fingers, branch to brush, with ash, coal, powder and/or paste; (P3) *other hygiene practice*: shoe-wearing habits at school; shoe-wearing habits at home (inside latrine and in courtyard), plus P1and P2; (P4) *hand cleanliness*: clean hands, trimmed nails, and clean nails; (P4E) *hand cleanliness + adenosine triphosphate (ATP)*: ATP improvement rate comparing before and after handwashing (detailed below, taken for 10% of child samples for the endline only due to invalid measurement in the baseline); (P5) *nutrition practice*: breakfast taken, and food taken for the past three days, ordered by the richness of the nutrition score and categorised into seven categories (none; carbohydrate (and fat); carbohydrate and vitamins; vegetable/animal protein and vitamins; vegetable/animal protein and carbohydrate; vegetable protein, animal protein and carbohydrate; protein, carbohydrate and vitamins); (P6) *health/hygiene knowledge*: handwashing procedure; breakfast important; (P6E) *health/hygiene knowledge + extra*: P6 outcomes plus putting water in latrine before defecating (endline only), oral rehydration solution (ORS) making (endline only), and food pyramid (endline only). Although it is not

easy to measure KAPB, handwashing evaluated pupils' demonstration of handwashing procedures. Also, *hand cleanliness* was not only checked by the surveyors but also measured via ATP which measures actively growing microorganisms through the light produced through their reaction with the naturally occurring firefly enzyme luciferase using a luminometer. Although smaller figures suggest cleaner hands, i.e., less microorganisms, there are large individual variations in the figures. Thus, tests were conducted before and after handwashing to measure whether appropriate handwashing had been done. The ATP improvement rate was calculated as (ATP-before – ATP-after)/ATP-before; thus, the larger the figure the higher the improvement.

The secondary, more indirect outcomes concern child health improvements as follows: (I1) *anthropometry*: height-, weight- (net of clothes) and BMI-for-age-z-score; (I2) *cold-related symptoms*: cough, breathing difficulty, sore throat, fever, running nose, and congested nose at present and in the past two-weeks; (I3) *other illness*: diarrhoea, stomachache, scabies in the past two-weeks, fatigue, dizziness, and appetite loss. Indicators that require two-week recall might entail recall bias, although no difference was expected across treatment groups. Apart from disease in the past two-weeks which was report-based, children's health status and outer symptoms were directly checked by the surveyors.

As for additional controls, child household wealth and parents' literacy level are considered, since family wealth and education level may affect child behaviour acquisition despite a school-based intervention directly targeting the children may mitigate the family socioeconomic influence.

Child-level empirical model

For child-level analysis, it is essential to take care of non-random differences in mean values at the baseline for several key outcome variables. The equivalent of DID equation (2) above, adding $\mathbf{Z}_{ijt}$, a vector of exogenous variables for child *i* in school *j* at period *t* is:

(3) $Y_{ijt} = \alpha + \beta \cdot t \cdot T_j + \delta \cdot t + \gamma \cdot T_j + \varphi \cdot \mathbf{X}_{jt} + \theta \cdot \mathbf{Z}_{ijt} + \theta D_j + v_{ijt}$,

A composite error is $v_{ijt} = c_j + \omega_{ij} + u_{ijt}$, where c_j is unobserved group heterogeneity, ω_{ij} is unobserved child heterogeneity, and u_{ijt} is the idiosyncratic error. Additionally, given that the RCT was executed at the school level, child-level estimations need to take into account possible intracluster correlation within school, defined as the ratio of within-cluster variance divided by total variance, which is the sum of within-cluster and between-cluster variance. The usual practice of removing the effects of intracluster correlation (ρ) is to conduct the analysis with intervention-unit clustering, adjusting for standard errors (SE) in regressions. With the hierarchical structure of school, time and child, we apply the most commonly and widely used DID estimation with multilevel random-effects (RE) with cluster-robust SE (CRSE) [40–42]. However, CRSE is not a panacea [38,42] that especially with a single baseline and follow-up, it would require twice the sample size in DID to get the same power as obtained with the analysis of covariance (ANCOVA) [43]. Also, in case the treatment effect extends beyond the individual level and affects the entire cluster, removing the intracluster correlation can underestimate the true effect. We therefore conduct SUR for family-wise outcomes with and without CRSE. For single outcome estimation, we also apply ANCOVA estimation and its likewise alternative constrained baseline analysis (CBA) which are reported to be more efficient than DID-CRSE or post estimator which uses only the endline data. Detailed description of the model and the estimated treatment effects and interpretation for child-level single outcomes are given in the Supporting Information (A-Table 5).

Exogenous covariates considered in the empirical models are child *sex*, *age*, *child wealth index* and *parents' literacy*. *Child wealth index* is created through iterated principal factor, reflecting house structure materials, roof materials, number of rooms, latrine structure and materials, possession of electronic appliances, mobile phones and bikes. The above-cited and another study [44], which additionally provided households with drinking water disinfection, showed that households in the squatter settlement with lower socioeconomic status took longer—several months—to acquire a healthy habit. On the other hand, another study [45] showed that well-specific arsenic information

induced costly behaviour of well change, with no influence of household wealth level although with positive influence of education level.

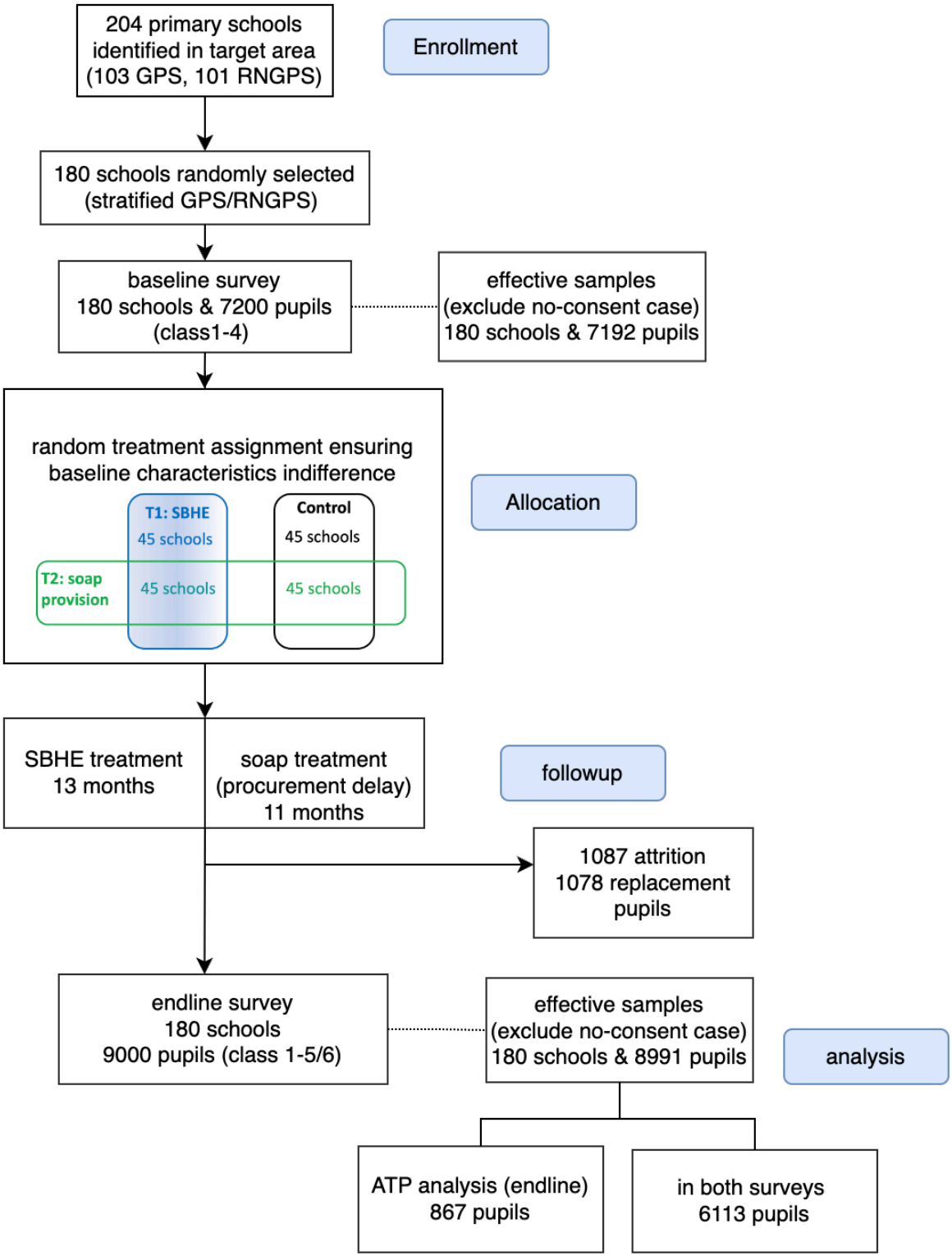
Given the fact that 15% of children in the endline survey were replacement of those attritted, we conduct estimations also for the subset of children who were in both baseline and endline.

Results

Project implementation

The baseline survey was conducted before the intervention, with participants recruitment occurring from 13 October–29 November 2011, mostly during cooler winter season, and the endline data were collected during 7 April–6 July 2013, mostly during hot and humid summer, after the completion of the intervention (March 2012–March 2013). The planned intervention period was from January to November 2012, based on the academic year in Bangladesh. However, the start of the implementation was shifted back for a few months due to a procedural delay. The randomisation was done in January 2012 after the beginning of academic year in Bangladesh, thus the treatment status should not have affected the choice of school by the children and their families. Treatment randomisation ensured statistical nondifference between treatment and control schools at school level variables. We collected data from 180 schools, with 45 each randomly assigned for a cross-cutting HE-SP treatments. The project profile depicting the participant flow with randomisation design is provided in Figure 1.

Figure 1. Study Profile with Participant Flow and the Randomisation Scheme



All HE-treatment schools were delivered all seven SBHE modules by the para-teachers. They also guided a cleaning rota by pupils. There were a few extra weeks as well as several school health events in which skill-trainings for the first two modules and illness treatment, such as ORS making, were repeated. As part of the intervention, the treatment group pupils were encouraged to measure their weight and height occasionally, which was expected to increase their health awareness. The SP intervention was not implemented as planned due to delayed procurement for a few months, resulting in irregular distribution. The SP schools were not monitored in terms of how they used the distributed soap bars, and 12 out of 45 SP schools and 8 out of 45 HESP schools reported not receiving soap, although the target pupils/households of those soap treatment schools were reported to have received the soap, suggesting possible recording errors at the school level. While the comparison of HE-treatment and SP-treatment schools was anticipated to provide insight into the direct effect of SBHE, independent of Hawthorn effects, as well as to discern the effect of mere goods provision on promoting healthy behaviour, the imperfect implementation of the SP-treatment raised certain concerns about the validity of this attempt. Thus, the cross-cutting intervention results are mainly contained in the Supporting Information, with limited interpretation here. The statistical analysis results below focus on the HE-treatment effect.

A refresher workshop was conducted in the midway for their experience-sharing, feed-back and suggestions by para-teachers and project staff. Some feedbacks were: difficulty in teaching micro-nutrients; need of additional materials to explain vitamin deficiency; need of shades/curtains for the window in order to use the projector; problems with locked/unfunctional latrines; use of projector being beneficial; students liking cartoons; children asking mothers to wash hands properly; children asking mothers to cut their nails; making toilet brush with palm tree leaves.

Child attrition

Between the baseline and endline, the study experienced a 15% attrition rate (1,087 out of 7,200 children), falling within the anticipated 20% attrition outlined in our sample size calculation. Specifically, both HE-treatment and HE-control groups saw attrition rates of 15.0%. While our focus is on HE-treatment effects, attrition rates were 13.7% and 16.3% for SP-treatment and SP-control groups, respectively, and 15.5%, 12.9%, 14.6%, and 17.1% for the HE-only, SP-only, HESP-treatment, and control schools, respectively. This could suggest that soap provision, though imperfectly implemented, may have affected a higher retention rate in SP-schools.

Most of the children in grade 5 during the intervention, thus grade 4 at baseline, had already finished primary school at the time of the endline survey, and their data were collected at the nearby high schools, while household follow-ups for those who did not go on to the neighbouring high schools proved to be mostly unsuccessful. For students who were in grade 4 at baseline and progressed to grade 5 during the intervention, endline data was gathered from 79.5% of them. Among these, 89.3% had transitioned to a nearby high school, while the remainder repeated the grade in primary school. The school-wide class repetition rate at the endline among our sample children amounted to 20.2%, of which 91.3% were single repetition. The repetition figure among our sample was much higher than official school figures of 9.4%. Attritted children mostly had gone to other regions. We collected replacement data for 15% or 1,078 pupils for those who attritted at the endline. The replacements were of the same sex and the same baseline class. With the addition of those who were in grade 1 at the time of intervention, the total effective sample size was 8,991 at the endline compared to 7,192 at the baseline (see Figure 1).

At the child level, there were some systematic differences between attritors and non-attritors suggested by a simple t-test of baseline characteristics. More male, older (and thus, other anthropometric measures), and unhealthy pupils—in particular, with more cold-related symptoms in two weeks, more diarrhoea symptoms, and more incidences of fatigue and appetite loss—attritted,

although no statistical difference was seen between healthy behaviour and practices. The reasons for attrition could not be retrieved. Regressing attrition on the baseline variables, each interacted with HE treatment, HE-schools had higher rate of attritors experiencing diarrhoea, while HE-control schools had higher rate of attritor experiencing stomachache. No other statistically significant difference was observed between HE-treatment and HE-control schools. This suggests that no clear one direction treatment bias is expected due to attrition. Thus, if these attrition tendencies affected the evaluation outcome, it would be a downward bias rather than an upward bias if the treatment exerted a positive impact, especially on less healthy and/or older pupils.

General baseline observations

Descriptive statistics for the main variables at the baseline for HE-treatment and HE-control groups as well as cross-cutting HESP-treatment groups are provided below. Outcome variables are categorised into families as explained above. Below we provide general observations at baseline.

School-level data

Table 2 gives descriptive statistics for the baseline school characteristics and outcomes variables categorised in their analytic families for (1) HE-treatment groups (non-HE and HE) and for (2) cross-cutting HESP-treatment groups (control, SP, HE and HESP). Although school type, GPS/RNGPS is not specified in the table, female teacher numbers and ratios are 3.1 (1.72) and 0.58 (0.27) for GPS, and 1.01 (0.87) and 0.28 (0.25) for RNGPS, respectively (SD in parentheses). The randomisation ensured no statistical difference for school-level outcomes, namely, (S1) *school hygiene practice & maintenance*, (S2) *school hygiene infrastructure*, and (S3) *schooling*, and school baseline characteristics, apart from female teacher ratio. There was no baseline statistical difference among the HESP-treatment groups. In terms of (S1) *school hygiene practice & maintenance*, relatively low value for *routine cleaning* was noticeable, being less than once a week, although there were high variabilities

across schools. Relatively high variabilities across schools were also suggested for *clean latrine*, *rubbish bin provision*, and *rubbish treatment*. Latrine availability was an important issue in Bangladeshi primary schools, that they were often lacking in numbers, unclean and/or locked up. For (S2) *school hygiene infrastructure*, the availability of usable latrines was notably low in both HE-control and HE-treatment schools, with an average of 0.34 and 0.39 clean, unlocked, pit-sealed latrines for boys, and 0.34 and 0.41 for girls, respectively. Considering that most grades had half-day school sessions, there were, on average, 37 boys and 41 girls attending school per each unlocked and clean latrine. Regarding (S3) *schooling*, the absenteeism rate differed across grades, ranging from 33% in grade 1 to 12% in grade 5, while the average number of students was highest in grade 1, with 52.7 students, and lowest in grade 5, with 31 students. *Dropout rate*, available except for grade 1, was not high but increasing in grade, from an average rate of 0.05% in grade 2 to an average rate of 0.85% in and grade 5. *Repetition rate*, available except for grade 1, showed its highest value for grade 3 with its mean value at 9.5%, while it is the lowest for grade 5 at 1.1%.

Table 2. Summary Statistics of Baseline School Characteristics by the HE-treatment and the Cross-cutting HESP-treatment

	HE-intervention			Cross-cutting HESP-intervention				
	non-HE school	HE school		control school	SP-school	HE-school	HESP-school	
	Mean (SD)	Mean (SD)	p-value [(a)]	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	p-value [(a)]
number of schools	90	90		45	45	45	45	
total number of enrolled pupils	232.4 (95.7)	220.8 (97.5)	0.420	229.2 (89.0)	235.6 (102.9)	219.9 (108.3)	221.7 (86.5)	0.860
total number of attended pupils	164.9 (73.3)	159.8 (71.2)	0.640	157.4 (68.0)	172.4 (78.2)	157.4 (77.8)	162.2 (64.7)	0.730
ratio of female pupils enrolled	0.51 (0.04)	0.50 (0.05)	0.525	0.50 (0.04)	0.51 (0.04)	0.50 (0.04)	0.51 (0.05)	0.453
ratio of female pupils attended	0.53 (0.05)	0.53 (0.05)	0.529	0.53 (0.05)	0.53 (0.06)	0.52 (0.04)	0.54 (0.06)	0.357
ratio of female teachers	0.48 (0.31)	0.38 (0.28)	0.026	0.49 (0.31)	0.47 (0.31)	0.37 (0.25)	0.39 (0.31)	0.163
Primary Outcomes by Families								
S1: school hygiene & maintenance								
routine latrine cleaning	0.37 (0.48)	0.38 (0.49)	0.878	0.38 (0.49)	0.36 (0.48)	0.40 (0.50)	0.36 (0.48)	0.968

latrine cleaning days/week	0.61 (1.09)	0.66 (1.17)	0.792	0.58 (0.92)	0.64 (1.25)	0.73 (1.27)	0.58 (1.08)	0.903
routine classroom cleaning	0.80 (0.40)	0.72 (0.45)	0.223	0.84 (0.37)	0.76 (0.43)	0.82 (0.39)	0.62 (0.49)	0.059
classroom cleaning days/week	3.48 (2.49)	2.98 (2.59)	0.188	3.47 (2.48)	3.49 (2.53)	3.24 (2.46)	2.71 (2.71)	0.439
classroom rubbish bin provision	0.29 (0.71)	0.20 (0.60)	0.365	0.31 (0.73)	0.27 (0.69)	0.22 (0.64)	0.18 (0.58)	0.797
latrine brush provision	0.63 (0.49)	0.61 (0.49)	0.804	0.57 (0.50)	0.69 (0.47)	0.69 (0.47)	0.53 (0.50)	0.294
rubbish treatment[(b)]	0.40 (0.86)	0.42 (0.86)	0.863	0.40 (0.89)	0.40 (0.84)	0.49 (0.94)	0.36 (0.77)	0.904
clean classrooms	1.19 (0.89)	1.00 (0.91)	0.160	1.07 (0.91)	1.31 (0.85)	0.96 (0.90)	1.04 (0.93)	0.279
soap for pupil at handwashing facility	0.46 (0.80)	0.40 (0.75)	0.630	0.44 (0.78)	0.47 (0.81)	0.47 (0.81)	0.33 (0.67)	0.821
Secondary Outcomes by Families								
S2: school hygiene infrastructure								
usable unlocked clean pit-sealed latrine for boys[(c)]	0.39 (0.63)	0.34 (0.56)	0.619	0.38 (0.53)	0.40 (0.72)	0.27 (0.54)	0.42 (0.58)	0.617
usable unlocked clean pit-sealed latrine for girls[(c)]	0.41 (0.65)	0.34 (0.62)	0.483	0.40 (0.58)	0.42 (0.72)	0.31 (0.60)	0.38 (0.65)	0.859
ratio of clean-usable latrines for boys[(c)]	0.47 (0.48)	0.51 (0.47)	0.567	0.51 (0.47)	0.43 (0.49)	0.49 (0.47)	0.53 (0.47)	0.764
ratio of clean-usable latrines for girls[(c)]	0.49 (0.48)	0.46 (0.47)	0.675	0.52 (0.48)	0.46 (0.49)	0.44 (0.47)	0.48 (0.47)	0.87
school hand washing facility near latrine	0.68 (0.47)	0.71 (0.46)	0.63	0.76 (0.43)	0.60 (0.50)	0.64 (0.48)	0.78 (0.42)	0.199
youngest children can get water by themselves	0.62 (0.49)	0.63 (0.48)	0.878	0.56 (0.50)	0.69 (0.47)	0.67 (0.48)	0.60 (0.50)	0.545
S3: schooling								
grade1 absence rate	0.35 (0.18)	0.31 (0.17)	0.135	0.38 (0.20)	0.32 (0.16)	0.32 (0.17)	0.30 (0.17)	0.154
grade2 absence rate	0.30 (0.18)	0.28 (0.15)	0.577	0.33 (0.21)	0.27 (0.14)	0.31 (0.16)	0.26 (0.15)	0.127
grade3 absence rate	0.30 (0.18)	0.31 (0.16)	0.964	0.33 (0.20)	0.27 (0.16)	0.31 (0.15)	0.30 (0.17)	0.411
grade4 absence rate	0.30 (0.17)	0.29 (0.15)	0.415	0.31 (0.17)	0.29 (0.16)	0.29 (0.15)	0.28 (0.14)	0.79
grade5 absence rate	0.12 (0.12)	0.12 (0.14)	0.852	0.12 (0.12)	0.13 (0.12)	0.15 (0.16)	0.09 (0.11)	0.323
grade2 dropout rate	0.07 (0.45)	0.04 (0.33)	0.666	0.00 (0.00)	0.14 (0.63)	0.00 (0.00)	0.08 (0.46)	0.276
grade3 dropout rate	0.16 (1.10)	0.12 (0.65)	0.739	0.15 (0.99)	0.18 (1.20)	0.00 (0.00)	0.24 (0.91)	0.64
grade4 dropout rate	0.45 (1.52)	0.40 (1.39)	0.82	0.58 (1.71)	0.32 (1.31)	0.51 (1.71)	0.30 (0.99)	0.74
grade5 dropout rate	0.80 (2.47)	0.89 (2.42)	0.805	0.85 (2.94)	0.76 (1.93)	1.00 (2.66)	0.79 (2.17)	0.966
grade2 repetition rate	6.62 (8.71)	6.00 (7.26)	0.606	7.52 (10.15)	5.71 (6.98)	6.55 (7.77)	5.45 (6.77)	0.608
grade3 repetition rate	9.54 (10.13)	9.40 (9.74)	0.923	10.06 (9.56)	9.03 (10.76)	8.56 (9.98)	10.23 (9.53)	0.83
grade4 repetition rate	9.37 (13.23)	8.80 (10.68)	0.753	10.26 (11.18)	8.48 (15.07)	8.01 (9.97)	9.60 (11.40)	0.806
grade5 repetition rate	0.94 (2.82)	1.21 (2.71)	0.512	0.72 (2.34)	1.16 (3.23)	0.77 (1.78)	1.66 (3.36)	0.343

Notes: (a) t-test for numerical variables and chi-square test for categorical variables; (b) ranging from 0 none, 1 dump in river/pond, 2 dump in another land, 3 burn/burry in school; (c) includes unisex toilets.

Child-level data

For child-level variables in Table 3, base characteristics shown are child *sex*, *age*, *child wealth index* and *parents' literacy*. While randomisation assured statistical indifference across school-level variables, it was not possible to do so across various child-level variables. Since we focus on HE-treatment, those variables with statistical difference at the 5% statistical significance level among the HE-treatment groups are indicated with an asterisk in the p-value column. Such non-random baseline differences are dealt by applying the DID estimator. For the primary child-level outcomes concerning health-related KAPB change, non-HE (control) group exhibited higher mean values for most outcomes in the (P1) *handwashing practice* family. Nonetheless, one of the outcomes in the (P4) *hand cleanliness*, *clean hands*, which could be indicative of their proper *handwashing practice* was more frequent in the HE-children. Among other healthy KAPB outcomes, *shoes/footwear wearing at school* in (P3) *overall hygiene* had higher mean value for non-HE children, while HE children have slightly higher mean scores for (P5) *nutrition practice*.

Table 3. Summary Statistics of Baseline Child Characteristics by the HE-treatment and the Cross-cutting HESP-treatment

	HE-intervention			**Cross-cutting HESP-intervention**				
	non-HE school	HE school		control school	SP-school	HE-school	HESP-school	
	Mean (SD)	Mean (SD)	p-value [(a)]	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	p-value [(a)]
number of surveyed children	3,594	3,598		1,797	1,797	1,799	1,799	
sex (male:1; female:2)	9.15 (1.59)	9.19 (1.60)	0.369	9.07 (1.58)	9.24 (1.61)	9.21 (1.60)	9.17 (1.59)	0.437
age (year)	9.15 (1.59)	9.19 (1.60)	0.369	9.07 (1.58)	9.24 (1.61)	9.21 (1.60)	9.17 (1.59)	0.007
child (family) wealth index[(b)]	-0.09 (0.87)	-0.13 (0.83)	0.056	-0.06 (0.91)	-0.13 (0.84)	-0.14 (0.82)	-0.13 (0.83)	0.037
parents' literacy (read)	1.03 (0.85)	1.02 (0.86)	0.688	1.05 (0.84)	1.02 (0.86)	1.05 (0.86)	1.00 (0.86)	0.252
parents' literacy (write)	1.02 (0.85)	1.01 (0.86)	0.508	1.04 (0.84)	1.01 (0.86)	1.04 (0.86)	0.98 (0.86)	0.113

Primary Outcomes by Families

P1: handwashing practice								
handwashing with soap before eating	0.19 (0.39)	0.14 (0.35)	<0.001*	0.20 (0.40)	0.19 (0.39)	0.14 (0.35)	0.15 (0.36)	<0.001
handwashing with soap after defecation	0.65 (0.48)	0.62 (0.48)	0.03*	0.65 (0.48)	0.65 (0.48)	0.63 (0.48)	0.62 (0.48)	0.19
handwashing with soap after playing	0.09 (0.29)	0.07 (0.25)	<0.001*	0.10 (0.30)	0.08 (0.27)	0.06 (0.23)	0.08 (0.27)	<0.001
handwashing index before eating[(c)]	3.17 (1.49)	3.06 (1.36)	0.002*	3.15 (1.52)	3.19 (1.45)	3.10 (1.31)	3.03 (1.40)	0.006
handwashing index after defecation[(c)]	4.79 (1.31)	4.72 (1.30)	0.031*	4.74 (1.33)	4.84 (1.29)	4.73 (1.28)	4.71 (1.33)	0.016
handwashing index after playing[(c)]	2.25 (1.46)	2.27 (1.37)	0.4	2.20 (1.47)	2.29 (1.44)	2.38 (1.32)	2.17 (1.40)	<0.001
handwashing under running water	0.99 (0.22)	0.97 (0.22)	0.011*	0.98 (0.22)	0.99 (0.21)	0.97 (0.23)	0.98 (0.21)	0.033
handwashing procedures[(d)]	1.41 (0.80)	1.38 (0.85)	0.101	1.43 (0.79)	1.40 (0.80)	1.42 (0.85)	1.34 (0.84)	0.015
P2: dental care practice								
dentalcare index[(e)]	9.30 (5.04)	9.17 (4.78)	0.26	9.52 (5.03)	9.08 (5.05)	8.97 (4.81)	9.36 (4.74)	0.002
tooth brushing frequency	1.49 (0.97)	1.47 (0.93)	0.214	1.54 (0.95)	1.45 (0.97)	1.43 (0.94)	1.50 (0.92)	0.002
dental care materials[(f)]	4.53 (2.25)	4.56 (2.24)	0.683	4.63 (2.20)	4.44 (2.30)	4.46 (2.28)	4.65 (2.20)	0.007
P3: overall hygiene practice (+ P1~P3)								
wearing shoes at school	0.88 (0.33)	0.85 (0.35)	0.002*	0.87 (0.34)	0.89 (0.32)	0.85 (0.36)	0.86 (0.34)	0.002
wearing shoes at home (latrine + yard)	2.97 (1.11)	2.92 (1.14)	0.056	2.97 (1.13)	2.98 (1.09)	2.89 (1.16)	2.96 (1.13)	0.078
P4: hand cleanliness								
clean hands	0.48 (0.50)	0.51 (0.50)	0.02*	0.46 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.014
trimmed nails	0.79 (0.92)	0.81 (0.93)	0.333	0.82 (0.93)	0.76 (0.91)	0.84 (0.94)	0.78 (0.92)	0.033
clean nails	0.57 (0.85)	0.58 (0.85)	0.338	0.59 (0.86)	0.54 (0.83)	0.59 (0.86)	0.58 (0.85)	0.234
P5: nutrition practice								
breakfast habit	1.89 (0.34)	1.91 (0.30)	0.008*	1.89 (0.34)	1.89 (0.35)	1.92 (0.29)	1.90 (0.31)	0.022
breakfast eaten today[(g)]	1.54 (0.53)	1.57 (0.52)	0.017*	1.55 (0.53)	1.53 (0.53)	1.60 (0.51)	1.55 (0.52)	0.001
food taken in 3 days[(g)]	1.93 (0.26)	1.94 (0.23)	0.004*	1.93 (0.25)	1.92 (0.27)	1.94 (0.24)	1.95 (0.22)	0.015
P6: health knowledge								
breakfast important	0.59 (0.49)	0.54 (0.50)	<0.001*	0.57 (0.49)	0.60 (0.49)	0.54 (0.50)	0.54 (0.50)	<0.001
handwashing procedures	1.41 (0.80)	1.38 (0.85)	0.101	1.43 (0.79)	1.40 (0.80)	1.42 (0.85)	1.34 (0.84)	0.015
Secondary Outcomes by Families								
I1: anthropometry[(h)]								
height z-score	-1.58 (1.23)	-1.57 (1.20)	0.632	-1.51 (1.23)	-1.66 (1.22)	-1.57 (1.23)	-1.57 (1.17)	0.004
weight z-score (net of clothes)	-0.84 (1.24)	-0.83 (1.23)	0.633	-0.76 (1.26)	-0.92 (1.22)	-0.86 (1.22)	-0.79 (1.24)	<0.001
BMI z-score	-1.56 (1.06)	-1.55 (1.02)	0.706	-1.53 (1.08)	-1.59 (1.04)	-1.52 (1.05)	-1.58 (0.98)	0.074
I2: cold-related symptoms								

cough	0.22 (0.42)	0.23 (0.42)	0.533	0.23 (0.42)	0.22 (0.41)	0.22 (0.42)	0.23 (0.42)	0.685
breathing difficulty	0.01 (0.10)	0.01 (0.11)	0.394	0.01 (0.10)	0.01 (0.11)	0.01 (0.10)	0.02 (0.13)	0.248
sore throat	0.01 (0.07)	0.01 (0.08)	0.297	0.01 (0.09)	0.00 (0.05)	0.01 (0.07)	0.01 (0.10)	0.056
fever	0.07 (0.25)	0.08 (0.27)	0.041*	0.07 (0.25)	0.07 (0.26)	0.08 (0.28)	0.08 (0.27)	0.17
running nose	0.31 (0.46)	0.32 (0.47)	0.142	0.30 (0.46)	0.31 (0.46)	0.32 (0.47)	0.32 (0.47)	0.464
congested nose	0.05 (0.22)	0.04 (0.21)	0.238	0.05 (0.22)	0.05 (0.21)	0.04 (0.20)	0.05 (0.21)	0.559
cough within 2 weeks	0.38 (0.49)	0.41 (0.49)	0.027*	0.39 (0.49)	0.38 (0.49)	0.41 (0.49)	0.41 (0.49)	0.176
breathing difficulty within 2 weeks	0.02 (0.14)	0.03 (0.16)	0.114	0.02 (0.15)	0.02 (0.12)	0.02 (0.14)	0.03 (0.17)	0.029
sore throat within 2 weeks	0.02 (0.13)	0.02 (0.15)	0.094	0.02 (0.14)	0.02 (0.12)	0.02 (0.13)	0.03 (0.17)	0.008
fever within 2 weeks	0.25 (0.43)	0.26 (0.44)	0.331	0.23 (0.42)	0.27 (0.44)	0.27 (0.44)	0.25 (0.43)	0.011
running nose within 2 weeks	0.48 (0.50)	0.51 (0.50)	0.002*	0.47 (0.50)	0.48 (0.50)	0.52 (0.50)	0.51 (0.50)	0.014
congested nose within 2 weeks	0.09 (0.29)	0.10 (0.29)	0.5	0.10 (0.30)	0.08 (0.27)	0.09 (0.28)	0.10 (0.31)	0.059
I3: infectious disease (+I2)								
diarrhoea within 2 weeks	0.15 (0.36)	0.16 (0.36)	0.527	0.15 (0.36)	0.15 (0.36)	0.17 (0.37)	0.15 (0.36)	0.46
stomachache within 2 weeks	0.36 (0.48)	0.37 (0.48)	0.765	0.40 (0.49)	0.33 (0.47)	0.37 (0.48)	0.36 (0.48)	<0.001
scabies	0.14 (0.35)	0.13 (0.34)	0.12	0.15 (0.36)	0.14 (0.34)	0.12 (0.33)	0.14 (0.34)	0.185
I4: overall health (+I2, I3)								
fatigue	0.33 (0.47)	0.31 (0.46)	0.09	0.34 (0.47)	0.31 (0.46)	0.31 (0.46)	0.31 (0.46)	0.161
dizziness	0.36 (0.48)	0.35 (0.48)	0.335	0.36 (0.48)	0.35 (0.48)	0.35 (0.48)	0.34 (0.47)	0.398
appetite loss	0.27 (0.44)	0.27 (0.44)	0.617	0.27 (0.44)	0.27 (0.45)	0.26 (0.44)	0.27 (0.45)	0.66

Notes: (a) t-test for numerical variables and chi-square test for categorical variables; (b) child family wealth index is created through iterated principal factor, reflecting house structure materials, roof materials, number of rooms, latrine structure and materials, possession of electronic appliances, phones and bikes; (c) index reflects handwashing frequencies on each occasion and usage of soap, ash, mud, only water on each occasion, ordered [0, 6]; (d) correct handwashing procedure [0, 8]; (e) index reflects frequency of brushing (never, sometimes, always; once a day, twice a day) and use of brush/branch finger, use of toothpaste, ash, or nothing; (f) dental care material combinations (fingers, branch to brush, with ash, coal, powder and/or paste) ordered [0, 6]; (g) ordered by the richness of the nutrition score (none; carbohydrate (and fat); carbohydrate and vitamins; vegetable/animal protein and vitamins; vegetable/animal protein and carbohydrate; vegetable protein, animal protein and carbohydrate; protein, carbohydrate and vitamins); (h) 5, 2, 14 outliers omitted for weight, height, BMI anthropometric indicator, respectively; significant differences between no-HE and HE schools found by regressing them on the HE dummy are noted with asterisk (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

As for the health-related outcomes, (I1) *anthropometric measurements* all showed negative values, indicating Bangladeshi children's below international-mean values on average by a standard deviation (SD) of 1.58, 0.84 and 1.56, respectively. Note that recording accurate child age was not a straightforward task in developing countries, where records were often unavailable. The uncertainty

in determining an accurate age posed difficulty when calculating children's anthropometric z-scores. Four pieces of age information were gathered in the following order of priority: (1) birth certificate; (2) vaccination card; (3) mother's memory; (4) school register; (5) child's memory. In case of no official child-age record and sizable discrepancies among the figures in (3)~(5), height/weight measurements were also taken into accounts in determining the most probable age information. There were 79 cases for which no child age data was available, which were all at the baseline. Among the baseline and endline household survey data, accounting for 2,159 and 2,698 valid observations, respectively, 61% and 71% possessed either child's birth certificates or immunization cards; however, about 7% of those who possessed the certificate still could not remember the birthdate of the child correctly. Given the fact that WHO Growth Chart for weight was only available for 5-10 years, while that for height and BMI was available for 5-19 years, the British 1990 Growth Charts available for 0-20 years were used to calculate the z-scores. There were 21 flagged outlier z-scores, set at +-6SD for weight and -6SD~+5SD for height and BMI. The standard flag criteria for the British formula was set at +-5SD, however, given the general thinness of Bangladeshi children and the majority of outliers having weights between -6SD~-5SD, it was set as +-6SD. There were 5 cases of weight, 2 cases of height, and 14 cases of BMI z-score outliers.

In terms of (I2) *cold-related symptoms*, 35% pupils reported having at least one symptom presently. Outer symptoms such as *cough*, *running/congested nose*, *fever*, *beathing difficulty*, *sore throat* and *scabies* were verified by the enumerators at the time of interview. Among cold-related symptoms, cough and running nose were most common, seen in about 20~30% plus children. These symptoms in the past two-weeks were reported by about 40~50% of the children, and fever by about 25%. In terms of other symptoms in the past two-weeks *(2w)*, *diarrhoea*, *stomachache*, *scabies*, *fatigue*, *dizziness* and *appetite loss* were experienced by 15.5%, 36.4%, 13.6%, 31.7%, 35.1% and 26.8% of children, respectively. Girls showed higher prevalence of *fatigue*, *dizziness* and *appetite loss* than boys with statistical significance at 0.1% or 1%. These observations seemed to suggest relatively unhealthy

children on average. Among health symptoms, *fever*, *cough (2w)* and *running nose (2w)* prevalence were higher among the HE school children with statistical significance. The average temperature in Dhaka was 24 and 29 Celsius at the baseline and endline, respectively. Temperature-wise, cold symptoms were expected to be more prevalent in the baseline survey, although a period dummy would take care of such period-wise differences. While the intracluster correlation ρ was assumed to be 0.058 at the time of sampling calculation, ρ for some outcome variables, such as infectious disease symptoms, was higher than the assumed figure.

Analysis results and interpretation: SBHE effects

School-level estimation results

Table 4 gives the average-effect estimates for each of family-wise outcome for the HE-intervention using the DID estimator. Only the primary outcome related to hygiene practices had positive HE-treatment effect with statistical significance at the 1% level. The effect of HE-treatment on (S1) *school hygiene practice & maintenance* was 0.36 [0.17–0.55] ($p<0.001$). The result was robust for the inclusion of additional control, *female teacher ratio*. The negative significant coefficient of *school type* in primary outcome estimations suggests that GPS, which were better funded and generally larger in scale than RNGPS, had higher level of school hygiene practice. No observed HE-treatment effect on the secondary outcomes could well be that overall hygiene infrastructure did not significantly change since no infrastructural intervention was provided, and as for schooling, the official record of schooling (absence, dropout and repetition, as well as the admitted number of pupils) might not have reflected the real situation. As a piece of evidence, the repetition figures provided by the school records was far less than what we found in our randomly chosen surveyed pupils. The total repetition rate in the endline reported by the schools was 9% for all grade 1~5, while it was 20% for our child data. The period effect was pronounced with high positive effects (i.e., negative β-coefficient for *illness*) with statistical

significance except for *schooling*. Statistically significant positive effects of *period* suggest that schools, regardless of treatment status, improved in these outcomes at the endline since the baseline.

Table 4. School-level analysis results: family-wise mean-standardised HE-treatment effect on five outcome families (N=360)

	primary outcomes		**secondary outcomes**			
	(S1) school hygiene practice & maintenance		**(S2) hygiene infrastructure**		**(S3) schooling**	
	β-coefficient [95%CI] [p-value]		β-coefficient [95%CI] [p-value]		β-coefficient [95%CI] [p-value]	
HE-treatment	0.37***	0.37***	-0.02	-0.02	0.09	0.09
	[0.17,0.57]	[0.17,0.57]	[-0.27,0.24]	[-0.27,0.23]	[-0.07,0.25]	[-0.07,0.25]
	[0.000]	[0.000]	[0.905]	[0.898]	[0.259]	[0.260]
HE-group	-0.08	-0.08	-0.01	0.02	-0.04	-0.05
	[-0.24,0.08]	[-0.24,0.08]	[-0.20,0.19]	[-0.18,0.21]	[-0.14,0.06]	[-0.15,0.06]
	[0.349]	[0.340]	[0.954]	[0.880]	[0.385]	[0.392]
period	0.60***	0.60***	0.53***	0.53***	-0.06	-0.06
	[0.47,0.74]	[0.47,0.74]	[0.34,0.71]	[0.34,0.71]	[-0.19,0.06]	[-0.19,0.06]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.338]	[0.338]
school type	-0.18***	-0.19**	-0.04	0.02	0.07+	0.07
	[-0.28,-0.08]	[-0.30,-0.07]	[-0.17,0.08]	[-0.13,0.16]	[-0.01,0.15]	[-0.02,0.16]
	[0.000]	[0.002]	[0.493]	[0.842]	[0.075]	[0.125]
female teacher ratio		-0.01		0.21		-0.01
		[-0.22,0.19]		[-0.06,0.47]		[-0.18,0.15]
		[0.896]		[0.127]		[0.897]
N	360	360	360	360	360	360

Note: Each column represents a separate regression on a family of outcomes estimated by seemingly unrelated regressions (SUR), estimated by a feasible generalised least squares (FGLS) estimator. β-coefficient is the mean standardised effect from SUR applying the difference-in-differences (DID) model, controlling for school type. An additional covariate is female teacher ratio. Outcome family compositions are: (S1) school hygiene practice: latrine cleaning rota, latrine cleaning days per week, classroom cleaning rota, classroom cleaning days per week, classrooms/corridors cleanness, rubbish bin provision, latrine brush provision, rubbish disposal method, soap provision at pupils' handwashing facility; (S2) school hygiene infrastructure: clean accessible pit-sealed latrines for boys; clean accessible pit-sealed latrines for girls, ratio of clean-usable latrine for boys, ratio of clean-usable latrine for girls, closeness of handwashing facility to the latrines (within 10~11 steps), easiness of obtaining water by younger children; (S3) schooling: absence rate for each grade1~5, repetition rate and dropout rate for each grade 2~5. Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence interval and p-value in brackets. 360 samples are 90 HE-schools and 90 control-schools in the baseline and endline.

As mentioned above, the results of HESP-intervention are given in the A-Table 1. To briefly summarise the equivalent estimates of cross-cutting HESP-treatments, the lone-HE-treatment had no effect on any outcome, while the HESP-treatment exerted positive effect on *school hygiene practice and maintenance* with similar magnitude to that of HE-treatment in Table 4 at the 1% significance level. No lone-SP-treatment effect was observed. The estimated HE-treatment effects for single school-level outcomes are given in the A-Table 2.

Child-level estimation results

The main results of average-effects model are presented in Table 5 for all children (*all*) and the subset of children who were in both baseline and endline surveys (*in both*). As in the school-level results, statistically significant positive HE-treatment effects were observed for primary outcomes related to hygiene practice and behaviour, namely, (P1) *handwashing* by 0.21 [0.12–0.30] (p<0.001), (P2) *dentalcare* by 0.17 [0.08–0.26], (p=0.002), (P3) *overall hygiene* by 0.22 [0.14–0.31] (p<0.001), (P6) *health/hygiene knowledge* 0.44 [0.33–0.55] (p<0.001), (P6E) *knowledge + extra* 0.20 [0.15–0.25] (p<0.001). These significant effects were observed across *all* and *in both* estimated with CRSE. Estimations without CRSE were highly similar. While (P4) *clean hands*, observed by the surveyor did not find any difference between treatment groups, the non-CRSE estimate for (P1E) *cleans hands + ATP* indicated possible improved cleanliness of hands by handwashing among HE pupils, although significant only at the 10% level. Note, however, that the number of pupils with ATP measurement was limited to 10%, one per class. For secondary outcomes, the HE-treatment had positive effect (denoted by a negative coefficient) on (H2) *cold-related symptoms* by -0.05 [-1.0–0.01] (p=0.086). The statistical significance was only at the 10% level for estimation with CRSE yet it was at the 1% level for estimation without CRSE, which could be due to the possibility of the treatment affecting the entire cluster especially for infectious disease symptoms. The average intracluster correlation for *cold-related symptoms* was reduced from the baseline of ρ=0.051 to the endline of ρ=0.023. This could suggest a positive change in the school-level dynamics related to cold symptoms, due to reduced spread or severity of cold symptoms within schools, better hygiene practices and increased awareness leading to fewer instances of cold symptoms being transmitted within the school, improved overall health environment within the clusters reducing the risk of cold symptom.

The selected results of cross-cutting HESP-intervention analysis are given in the A-Table 4 for models which HE-treatment had statistically significant effects in HE-treatment analysis. Briefly summarising the estimates, HE-only-treatment and HESP-treatment had positive effect with statistical

significance on these outcomes, namely, *handwashing dentalcare*, *overall hygiene*, and *health/hygiene knowledge*, while SP-only-treatment had no impact. HESP treatment had higher impact than HE-only treatment on *handwashing* and *overall hygiene*, indicating possible additional positive effect of soap availability. Particularly for *handwashing* and *overall hygiene*, the sum of HE-only and HESP-treatment effects were higher than that of HE-treatment in Table 5, while the magnitudes of coefficients were same or highly similar for other outcomes which were not considered to involve soap utilisation, i.e., *dentalcare* and *knowledge*.

Table 5. Child-level analysis results: family-wise mean-standardised HE-treatment effect on nine outcome families for all children and children in both surveys

	Primary Outcomes								
	(P1) handwashing			**(P2) dentalcare**			**(P3) overall hygiene**		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
	all	all (no CRSE)	in both	all	all (no CRSE)	in both	all	all (no CRSE)	in both
HE	0.21***	0.22***	0.23***	0.17***	0.17***	0.17***	0.22***	0.22***	0.24***
	[0.12, 0.30]	[0.18, 0.25]	[0.14, 0.31]	[0.08, 0.26]	[0.12, 0.22]	[0.07, 0.26]	[0.14, 0.31]	[0.19, 0.25]	[0.15, 0.32]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
HE-group	-0.05	-0.05***	-0.06	-0.03	-0.03	-0.03	-0.06+	-0.06***	-0.07*
	[-0.13, 0.03]	[-0.07, -0.03]	[-0.13, 0.01]	[-0.10, 0.05]	[-0.07, 0.01]	[-0.11, 0.04]	[-0.12, 0.00]	[-0.08, -0.04]	[-0.13, -0.01]
	[0.200]	[0.000]	[0.112]	[0.486]	[0.190]	[0.413]	[0.069]	[0.000]	[0.033]
period	0.64***	0.64***	0.67***	0.40***	0.40***	0.43***	0.53***	0.53***	0.57***
	[0.58, 0.71]	[0.61, 0.66]	[0.61, 0.73]	[0.33, 0.47]	[0.36, 0.43]	[0.36, 0.50]	[0.47, 0.59]	[0.51, 0.55]	[0.51, 0.63]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
school type	-0.08**	-0.07***	-0.08**	-0.10**	-0.10***	-0.10***	-0.08**	-0.08***	-0.09***
	[-0.13, -0.02]	[-0.09, -0.05]	[-0.13, -0.02]	[-0.16, -0.03]	[-0.12, -0.07]	[-0.16, -0.04]	[-0.13, -0.03]	[-0.10, -0.07]	[-0.14, -0.04]
	[0.008]	[0.000]	[0.004]	[0.002]	[0.000]	[0.001]	[0.001]	[0.000]	[0.000]
sex	0.10***	0.09***	0.09***	0.04**	0.04***	0.04**	0.08***	0.08***	0.07***
	[0.08, 0.12]	[0.07, 0.11]	[0.07, 0.11]	[0.01, 0.07]	[0.02, 0.07]	[0.01, 0.07]	[0.06, 0.09]	[0.06, 0.09]	[0.06, 0.09]
	[0.000]	[0.000]	[0.000]	[0.002]	[0.001]	[0.008]	[0.000]	[0.000]	[0.000]
N	16181	16181	12224	16183	16183	12226	16163	16163	12211

	(P4) clean hands			(P4E) clean hands + ATP (endline)			(P5) nutrition		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
	all	all (no CRSE)	in both	all	all (no CRSE)	in both	all	all (no CRSE)	in both
HE	-0.01	-0.01	-0.01	0.08	0.08+	0.08	-0.04	-0.04*	-0.05+
	[-0.11, 0.10]	[-0.06, 0.04]	[-0.12, 0.10]	[-0.02, 0.17]	[-0.01, 0.16]	[-0.03, 0.19]	[-0.10, 0.01]	[-0.08, -0.00]	[-0.10, 0.01]
	[0.906]	[0.802]	[0.871]	[0.122]	[0.076]	[0.175]	[0.122]	[0.030]	[0.080]
HE-group	0.03	0.03+	0.04				0.06*	0.06***	0.06**
	[-0.04, 0.11]	[-0.00, 0.07]	[-0.04, 0.11]				[0.01, 0.10]	[0.03, 0.08]	[0.02, 0.11]
	[0.412]	[0.086]	[0.368]				[0.012]	[0.000]	[0.007]
period	0.40***	0.40***	0.45***				-0.01	-0.01	0.007
	[0.31, 0.48]	[0.36, 0.43]	[0.36, 0.53]				[-0.05, 0.03]	[-0.04, 0.02]	[-0.03, 0.05]
	[0.000]	[0.000]	[0.000]				[0.604]	[0.475]	[0.725]
school type	-0.04	-0.04**	-0.04	0.01	0.01	0.03	-0.01	-0.01	-0.01
	[-0.10, 0.02]	[-0.07, -0.02]	[-0.10, 0.02]	[-0.09, 0.10]	[-0.08, 0.09]	[-0.08, 0.14]	[-0.04, 0.02]	[-0.03, 0.01]	[-0.04, 0.02]
	[0.184]	[0.002]	[0.173]	[0.870]	[0.851]	[0.620]	[0.423]	[0.252]	[0.387]
sex	0.15***	0.15***	0.16***	0.06	0.06	0.04	0.04***	0.04***	0.04**
	[0.12, 0.18]	[0.12, 0.17]	[0.13, 0.20]	[-0.02, 0.15]	[-0.02, 0.15]	[-0.05, 0.14]	[0.02, 0.06]	[0.02, 0.06]	[0.01, 0.06]
	[0.000]	[0.000]	[0.000]	[0.136]	[0.144]	[0.391]	[0.000]	[0.000]	[0.002]
N	16178	16178	12221	867	867	600	16183	16183	12226

	(P6) knowledge			(P6E) knowledge + extra (endline)		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
	all	all (no CRSE)	in both	all	all (no CRSE)	in both
HE	0.44***	0.44***	0.45***	0.20***	0.20***	0.21***
	[0.33, 0.55]	[0.39, 0.48]	[0.34, 0.56]	[0.15, 0.25]	[0.18, 0.22]	[0.15, 0.26]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
HE-group	-0.08	-0.08***	-0.06			
	[-0.18, 0.02]	[-0.11, -0.04]	[-0.16, 0.03]			
	[0.130]	[0.000]	[0.204]			
period	0.80***	0.80***	0.84***			
	[0.72, 0.88]	[0.77, 0.83]	[0.76, 0.92]			
	[0.000]	[0.000]	[0.000]			

school type	-0.02	-0.03*	-0.03	-0.04	-0.04**	-0.03
	[-0.09, 0.04]	[-0.05, -0.00]	[-0.10, 0.04]	[-0.09, 0.02]	[-0.06, -0.01]	[-0.09, 0.02]
	[0.433]	[0.022]	[0.375]	[0.171]	[0.002]	[0.244]
sex	0.03*	0.03*	0.02	0.03*	0.03**	0.02
	[0.00, 0.05]	[0.00, 0.05]	[-0.00, 0.05]	[0.01, 0.05]	[0.01, 0.05]	[-0.00, 0.05]
	[0.023]	[0.017]	[0.107]	[0.016]	[0.009]	[0.112]
N	16183	16183	12226	8991	8991	6113

Secondary Outcomes

	(H1) anthropometry			**(H2) cold-related symptoms**			**(H3) other illness**		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
	all	all (no CRSE)	in both	all	all (no CRSE)	in both	all	all (no CRSE)	in both
HE	-0.02	-0.02	-0.02	-0.05+	-0.05**	-0.05+	0.02	0.02	-0.004
	[-0.06, 0.02]	[-0.08, 0.03]	[-0.05, 0.01]	[-0.10, 0.01]	[-0.08, -0.01]	[-0.10, 0.01]	[-0.04, 0.08]	[-0.02, 0.05]	[-0.07, 0.06]
	[0.262]	[0.416]	[0.107]	[0.086]	[0.007]	[0.078]	[0.573]	[0.324]	[0.906]
HE-group	0.01	0.01	0.04	0.03	0.030*	0.029	-0.02	-0.02	-0.01
	[-0.06, 0.08]	[-0.03, 0.05]	[-0.03, 0.10]	[-0.01, 0.07]	[0.01, 0.05]	[-0.01, 0.07]	[-0.08, 0.04]	[-0.05, 0.01]	[-0.07, 0.05]
	[0.756]	[0.592]	[0.279]	[0.166]	[0.015]	[0.177]	[0.569]	[0.227]	[0.719]
period	0.14***	0.14***	0.18***	-0.05*	-0.05***	-0.05*	-0.19***	-0.19***	-0.16***
	[0.11, 0.17]	[0.10, 0.18]	[0.16, 0.20]	[-0.09, -0.00]	[-0.07, -0.02]	[-0.09, -0.00]	[-0.23, -0.14]	[-0.21, -0.16]	[-0.21, -0.11]
	[0.000]	[0.000]	[0.000]	[0.032]	[0.000]	[0.042]	[0.000]	[0.000]	[0.000]
school type	-0.01	-0.01	-0.003	-0.003	-0.003	-0.003	-0.01	-0.01	-0.01
	[-0.07, 0.06]	[-0.03, 0.02]	[-0.07, 0.06]	[-0.03, 0.03]	[-0.02, 0.01]	[-0.04, 0.03]	[-0.05, 0.03]	[-0.02, 0.01]	[-0.05, 0.03]
	[0.846]	[0.656]	[0.929]	[0.830]	[0.699]	[0.880]	[0.735]	[0.437]	[0.692]
sex	-0.05**	-0.05**	-0.04+	0.002	0.002	0.003	0.08***	0.08***	0.07***
	[-0.09, -0.01]	[-0.08, -0.02]	[-0.08, 0.00]	[-0.02, 0.02]	[-0.02, 0.02]	[-0.02, 0.02]	[0.06, 0.09]	[0.06 0.09]	[0.05, 0.09]
	[0.008]	[0.000]	[0.065]	[0.838]	[0.830]	[0.768]	[0.000]	[0.000]	[0.000]
N	16130	16130	12207	16183	16183	12226	16173	16173	12217

Notes: Each column represents a separate regression on a family of outcomes applying seemingly unrelated regressions (SUR), estimated by a feasible generalised least squares (FGLS) estimator with cluster-robust standard errors (CRSE) unless stated otherwise. β-coefficient is the mean-standardised effect from SUR applying the difference-in-differences (DID) model, controlling for school type and child sex. For each outcome family, estimates are provided for all sample children and children present *in both* baseline and endline. Each indicator family includes the following variables: (P1) *handwashing practice*: handwashing habits (washing frequency in each occasion (before eating, after defecation, after playing); used substances (soap, ash, mud and/or water only); washing with soap in each occasion; wash with running water; correct washing procedure); (P2) *dental care practice*: dentalcare (frequency and use of brush/branch, toothpaste, etc.), type of materials used; (P3) *overall hygiene practice*: shoes/footwear wearing at school (frequency), shoes/footwear wearing at home (frequency in latrine and in courtyard), + P1 & P2; (P4) *clean hand*s: clean hands by

observation, trimmed nails, clean nails; (P4E) *clean hand*s + *ATP*: additional hand cleanliness measured by ATP improvement rate (10% of samples); (P5) *nutrition practice*: breakfast habit, breakfast taken in 3 days, food taken in 3 days, ordered by the richness of nutrition score (none; carbohydrate (and fat); carbohydrate and vitamins; vegetable/animal protein and vitamins; vegetable/animal protein and carbohydrate; vegetable protein, animal protein and carbohydrate; protein, carbohydrate and vitamins); (P6) *health/hygiene knowledge*: handwashing procedure, breakfast significance; (P6E) *health/hygiene knowledge* + *extra:* additional knowledge measured only in the endline, i.e., putting water in latrine before defecating, oral rehydration solution (ORS) making, food pyramid; (H1) *anthropometry*: height-, weight-, BMI-z-score; (H2) *cold*: cold-related symptoms at present and in the past two-weeks, cough, breathing difficulty, sore throat, fever, running nose, congested nose; (H3) *other illness*: diarrhoea, stomachache, skin disease, fatigue, dizziness, appetite loss;. Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence intervals and p-value in brackets. The results are robust with additional adjustment with *child age*, *wealth index*, and *parent literacy* (see A-Table 3).

The above findings were robust to the exclusion of *sex*, as well as the inclusion of extra controls, *child age*, *child wealth index*, and *parents' literacy* which themselves often had statistically significant effects on the outcome families (estimation results with additional covariates in the A-Table 4). Estimated positive *sex* coefficients indicate that girls had better hygiene practice, yet negative coefficients for *anthropometry* and *other illness* indicate that girls had worse health status. Adding *child age* did not affect the HE-effect estimates but slightly reduced the magnitude of period effects. Child age had statistically significant positive effects on all primary outcomes while it had statistically significant negative effects on all secondary outcomes.

There was no sign of systematic baseline differences between treatment groups apart from *nutrition* and *overall hygiene* in which *HE-group* showed higher and lower baseline figures, respectively. *Period* was statistically significant at the 1% level for all estimation apart from *nutrition* model. Negative *period* coefficients in *cold-related symptoms* and *other illness* demonstrated improved health from baseline to endline. While the former could encompass seasonality effects as noted above, the latter was not obviously related to seasons. For other health-KAPB relevant outcomes, *period* exhibited positive effects with higher magnitude than *HE-treatment*. This signals general improvement in health-KAPB across pupils since the baseline, regardless of treatment status. Statistically significant negative effects of *school type* on *handwashing practice*, *dental-care practic*e and *other hygiene practice* suggested that pupils going to GPS school fared better in these aspects. Negative effects of school type on these outcome families indicate that RNGPS pupils had less of healthy practice. As a matter of fact,

not only the government school funding level but also the child household wealth indicator statistically differed between these school types. Thus, the type of school seems to be highly correlated to pupils' socioeconomic status. The estimated HE-treatment effects and interpretation for single child-level outcomes are given in the A-Table 5.

Period effect and of health education spillovers

Looking at the estimations results at both school and child levels, *period* had significant positive coefficient with large magnitudes regardless of treatment status for all outcomes but *schooling* and *child nutrition*. We can speculate several reasons for this: (1) there were general improvements in health practice and status across all schools due to external physical improvement such as hygiene infrastructure (but not general nutrition level); (2) there were certain HE-externality effects across schools in terms of skill and/or information transmission. While general health improvements are plausible along with general economic development, improvements in health KABP are suggestive of inter-school spillovers, possibly including John Henry effects among the control schools.

There were several factors that could have induced inter-school spillover effects. One was school proximities; schools were distributed in two sub-districts and their mean distance among two schools were 15 km, while minimum distance was 0.3 km and maximum was 41.4 km (see school location map in A-Figure 1). The average and maximum number of schools within 2 km, 3 km and 6 km of the HE schools were 1.7 and 6 schools, 3.9 and 10 schools, and 13.7 and 28 schools, respectively. In the area with several schools in proximity, it was not unusual for children from neighbouring households to go to different schools. In such circumstances, the skills and knowledge they had learned in the HE sessions might have been talked of. On the other hand, despite being in an era of communication technology, health skills which children learnt at schools would not be a likely topic discussed over mobile phones by their guardians, nor were such skills easily transmittable this way. Although the project implementation attempted to mitigate possible John Henry effects among non-treatment

schools as described above, such effects and consequent spillovers might have arisen especially given that HE-skills and knowledge were not complicated per se. Even for schools not in proximities, there were occasional school headteachers meetings and other casual teacher meetings where some conversation over this external project could have taken place. Thus, information transmission barrier could be low if sufficient interests existed, although practicing it and having behaviour change would be different. Examples of informational spillovers in field experiments abound [46–49]. We thus investigate whether there were any such spillovers, and if so, how they might have affected the treatment outcomes.

Empirical model with an externality term

We add a variable to capture possible inter-school spillover effects to the above estimation equations. Such externalities are captured by distances between schools, the number of schools in proximity, and by the number of pupils attending the school since HE-school pupils might have spoken about the HE contents to pupils from other schools. In the influential work by [8], inter-school externalities of deworming drug effects depended on the sums of the total number of students at school within a certain distance of a treatment school. In a recent work by [49], the spillover effects of agricultural advisory services were captured by the ratio of farmers who benefit via their social network.

Given that there is no clear theory or model construct as to how such information and skills/practices spread across distances and entities, the following externality variables are considered here: (1) a composite externality index which incorporates distances from the HE-treatment schools as well as the number of attending pupils; (2) the number of schools within certain radius of the HE-treatment schools (as well as in-between-distances). The composite externality index takes the following form: $\frac{1}{J \cdot \mu} \cdot \sum_k lnN_k^T \cdot lnN_j \cdot e^{-d_{kj}}$, where d_k is the distance of school j from a treatment school k, whose effect manifests exponential decay – the further the distance, the quicker the effects decay – multiplied by the natural logarithm of the total number of attending students in school j, lnN_j, and that in the treatment

school k, lnN_k, both measured at the baseline, summed up for all K treatment schools, and divided by total number of schools J times μ, the natural logarithm of the average total attending students of all J schools at the baseline, in order to normalise. Distances are calculated using global positioning system (*GPS*) data of each school location embedded in photos taken during the survey. For e^{-d}, the closer the schools, the lager the figure, and thus the spillover effect, e.g., for 500m the figure is 0.606, while for 3km and 6km, it is 0.050 and 0.002, respectively. We use baseline attending students' data rather than admitted students' data because the latter can be significantly different from the actual number of pupils. It was a routine practice for RNGPS schools to inflate the number of admitted students in order to obtain more school funds (a pairwise correlation between admitted and attended pupils for GPS and RNGPS at baseline were 0.933 and 0.819, respectively). It applies a log form because an increase in the number of contacts does not increase proportional to the number of pupils, given an increased probability of contacting the same person. Thus, incorporating the treatment externality index, the school-level DID estimation equation (2) becomes:

$$(5)\ Y_{jt} = \alpha + \beta \cdot t \cdot T_j + \delta \cdot t + \gamma \cdot T_j + \varphi \cdot \mathbf{X}_{jt} + \lambda \cdot \frac{1}{J \cdot \mu} \cdot \sum_k lnN_k^T \cdot lnN_j \cdot e^{-d_{kj}} + \theta D_j + \varepsilon_{jt}.$$

Here, λ is expected to reflect the magnitude of treatment externalities across schools, intensified by the number of attending students and closeness of schools both in a diminishing manner. The extent of externalities, that is, health-related information children learnt at the HE schools spreading beyond their families, thus depends on the number of students and the distance between the schools for the composite externality index.

Externality/spillover general observations

Considering the possibility of HE information dispersed by children, 2.6% of HE schools pupils reported to have communicated what they had learnt in SBHE with their school mates, 16.9% of them with their family members, 0.7% with pupils from other schools. Although children may have talked about the HE matters unconsciously and not remembering it, this observation suggests that

informational externality among children might not have been large. As noted, online information diffusion was neither likely among children nor parents. While there could be some informational spillover through family members, another possible route was at the school/teacher level through various meetings, particularly for the headteachers. It was quickly known that certain schools were receiving HE treatments, and the control schools were likely to have wished to be the next treatment target. Naturally, prior to the project commencement, we had to explain about the HE intervention to the district officer with the presence of several head teachers to obtain an agreement about the field experiment. If regular teachers induced spillovers, its effect could be partly captured by the stratification variable *school type* since GPS had higher number of teachers on average; the mean number of teachers for GPS and RNGPS at the baseline were 5.73 (1.48) and 3.72 (0.65), and of students are 198.0 (75.24) and 126.66 (47.0), respectively (SD in parentheses). A similar thing could be said regarding the female teachers' ratio. Note however, that teachers were not the direct target of SBHE and their active role was not necessarily expected. Below in Table 6, we provide a summary statics of externality measures used for the analysis, namely, *externality index*, and the number of schools *within1km*, *within 1-2km*, and *within 2-3km*.

Table 6. Summary statistics of externality measures (N=180)

	non-HE school N=90	HE Schools N=90	
	Mean (SD)	Mean (SD)	p-value [(a)]
HE externality index	0.44 (0.30)	0.39 (0.27)	0.254
number of HE schools within 1km	0.33 (0.56)	0.32 (0.52)	0.89
number of HE schools within 1-2km	1.42 (1.29)	1.22 (1.10)	0.264
number of HE schools within 2-3km	2.28 (1.50)	2.17 (1.41)	0.609

Notes: (a) t-test for numerical variables.

Externality model estimation results

We provide externality model estimation results for selected school- and child-level outcomes which had HE-treatment estimated with statistical significance in the original analysis.

School-level Spillover

The results of Table 7 indicate that there were likely to be positive inter-school spillover for the primary outcome, (S1) *school hygiene practice & maintenance*. The externality measures exhibited positive effect with statistical significance except for *within 2-3km*, while the significance, directions and magnitudes of *HE-treatment*, *period* effects as well as *school type* (omitted from Table 7) remained robust and unaffected. The *externality index* had the highest magnitude of effect, followed by the number of schools *within 1km, within 1-2km*, and *within 2-3km*, indicating larger spillovers among schools with more HE-schools in the nearby neighbourhood. The effect of *within 4km* or *within 3-4km* (neither shown here) was in the same direction with lesser magnitude. These results provide a strong support to the existence of positive HE externality in addition to the general period effects.

Table 7. Family-wise mean-standardised treatment effect and externality effect on the school-level primary outcome

	School-level Mean-standardized Treatment Effects with Externality Effects			
	(S1) school hygiene practice & maintenance			
	β-coefficient [95%CI] [p-value]	β-coefficient [95%CI] [p-value]	β-coefficient [95%CI] [p-value]	β-coefficient [95%CI] [p-value]
	externality measures			
	externality index	**within 1km**	**within 1-2km**	**within 2-3km**
HE- treatment	0.36***	0.36***	0.36***	0.36***
	[0.18, 0.55]	[0.18, 0.55]	[0.18, 0.55]	[0.17, 0.55]
	[0.000]	[0.000]	[0.000]	[0.000]
period	0.59***	0.59***	0.59***	0.59***
	[0.47, 0.72]	[0.46, 0.72]	[0.47, 0.72]	[0.46, 0.72]
	[0.000]	[0.000]	[0.000]	[0.000]
externality measures	0.27**	0.10*	0.05*	-0.01
	[0.11, 0.43]	[0.01, 0.20]	[0.01, 0.09]	[-0.04, 0.02]
	[0.001]	[0.025]	[0.020]	[0.596]
N	360	360	360	360

Notes: Each column represents a separate regression with different externality measures. Outcome families are same as in Table 4 primary outcomes. All estimations apply DID SUR, adjusted for *HE-group* and *school type* (estimates omitted). Externality index incorporate distances from all HE-treatment schools as well as the number of attending pupils $(1/(J \cdot \mu) \cdot \sum_k lnN_k^T \cdot lnN_j \cdot e^{-d_{kj}}$, where d_k is the distance of school j from a treatment school k, whose effect manifests exponential decay multiplied by the natural logarithm of the total number of attending students in school j, lnN_j, and that in the treatment school k, lnN_k, both measured at the baseline, summed up for all K treatment schools, and divided by total number of schools J times μ, the natural logarithm of the average total attending students of all J schools at the baseline, in order to normalize). Other externality measures ("within *x*-*y*km") correspond to the number of schools within

*x-y*km radius. Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence intervals and p-value in brackets. 360 samples are 90 HE schools and 90 HE-control schools in the baseline and endline.

Child-level Spillover

As in the school-level analysis, externality measures demonstrated significant impacts on (P1) *handwashing*, (P2) *dentalcare*, (P3) *overall hygiene*, (P6) *health/hygiene knowledge*, and (H2) *cold-related symptoms*. As in the school-level, *externality index* displayed the largest coefficient magnitude, followed by *within 1km*, *within 1-2km*, and *within 2-3km* (omitted here) among the externality measures (Table 8, estimates for *HE-group*, *school type* and *sex* are omitted). All estimated coefficients showed identical results as those in the original analysis given in Table 5. This not only reaffirms the robustness of *HE*, *period* and other estimates, but also indicates that externality effects were at work in addition to *HE-treatment* and *period* effects, and that they improved model's explanatory power. These statistically significant positive spillovers suggest that health-related information and possibly healthier behaviour spread to non-HE school pupils in proximity, with stronger effects for those having more of HE-schools closer by.

Table 8. Family-wise mean-standardized treatment effect and externality effect on the Child-level selected outcomes

	Child-level Mean-standardized Treatment Effects with Externality Effects								
	(P1) handwashing			**(P2) dentalcare**			**(P3) overall hygiene**		
	externality measures			**externality measures**			**externality measures**		
	externality index	**within 1km**	**within 1-2km**	**externality index**	**within 1km**	**within 1-2km**	**externality index**	**within 1km**	**within 1-2km**
HE-	0.21**	0.21**	0.21**	0.17**	0.17**	0.17**	0.22**	0.22**	0.22**
treatment	[0.13,0.30]	[0.13,0.30]	[0.13,0.30]	[0.08,0.26]	[0.08,0.26]	[0.08,0.26]	[0.14,0.31]	[0.14,0.31]	[0.14,0.31]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
period	0.64**	0.64**	0.64**	0.40**	0.40**	0.40**	0.53**	0.53**	0.53**
	[0.58,0.70]	[0.58,0.70]	[0.58,0.70]	[0.33,0.47]	[0.33,0.47]	[0.33,0.47]	[0.47,0.59]	[0.47,0.59]	[0.47,0.59]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
externality	0.18**	0.09**	0.04**	0.28**	0.10**	0.06**	0.23**	0.10**	0.04**
measures	[0.06,0.30]	[0.04,0.14]	[0.02,0.06]	[0.16,0.39]	[0.04,0.17]	[0.03,0.08]	[0.12,0.33]	[0.05,0.15]	[0.02,0.06]

	[0.003]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]
N	16181	16181	16181	16183	16183	16183	16163	16163	16163
	(P6) knowledge			**(P6E) knowledge + extra (endline)**			**(H2) cold-related symptoms (no CRSE)**		
	externality measures			**externality measures**			**externality measures**		
	externality index	**within 1km**	**within 1-2km**	**externality index**	**within 1km**	**within 1-2km**	**externality index**	**within 1km**	**within 1-2km**
HE-treatment	0.44***	0.44***	0.44***	0.20***	0.20***	0.20***	-0.05**	-0.05**	-0.05**
	[0.33,0.55]	[0.33,0.55]	[0.33,0.55]	[0.15,0.25]	[0.15,0.25]	[0.15,0.25]	[-0.08,-0.01]	[-0.08,-0.01]	[-0.08,-0.01]
	[0.000]	[0.000]	[0.000]				[0.007]	[0.007]	[0.007]
period	0.80***	0.80***	0.80***				-0.05***	-0.05***	-0.05***
	[0.72,0.88]	[0.72,0.88]	[0.72,0.88]				[-0.07,-0.02]	[-0.07,-0.02]	[-0.07,-0.02]
	[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]
externality measures	0.15*	0.07*	0.03*	0.10+	0.05+	0.02+	-0.05***	-0.04***	0
	[0.03,0.27]	[0.01,0.13]	[0.01,0.06]	[-0.01,0.21]	[-0.00,0.10]	[-0.00,0.04]	[-0.08,-0.02]	[-0.05,-0.02]	[-0.00,0.01]
	[0.018]	[0.021]	[0.016]	[0.089]	[0.056]	[0.081]	[0.001]	[0.000]	[0.526]
N	16183	16183	16183	8991	8991	8991	16183	16183	16183

Notes: Each column represents a separate regression with different externality measures. Outcome families are same as in Table 5 primary outcomes. All estimations apply DID SUR with CRSE, except for (H2) model, adjusted for *HE-group*, *school type*, and *sex* (estimates omitted). Externality index incorporate distances from all HE-treatment schools as well as the number of attending pupils $(1/(J \cdot \mu) \cdot \sum_k lnN_k^T \cdot lnN_j \cdot e^{-d_{kj}}$, where d_k is the distance of school j from a treatment school k, whose effect manifests exponential decay multiplied by the natural logarithm of the total number of attending students in school j, lnN_j, and that in the treatment school k, lnN_k, both measured at the baseline, summed up for all K treatment schools, and divided by total number of schools J times μ, the natural logarithm of the average total attending students of all J schools at the baseline, in order to normalize). Other externality measures ("within *x*-*y*km") correspond to the number of schools within *x*-*y*km radius. Significance level: +p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence intervals and p-value in brackets. 16183 samples belonged to 90 HE schools and 90 HE-control schools in the baseline and endline.

Cost-effectiveness of health education

The project's direct labour expenses for 18 para-teachers amounted to 14,466USD for the year. Additionally, there were costs for the initial training workshop and subsequent refresher workshops, totalling 1,156USD. Beyond these operational expenses, which summed up to 15,622USD, there were also fixed costs: 4,943USD for developing digital health education materials and 3,460USD for purchasing 23 mini-projectors, with five extras included. The HE digital materials can be used almost perpetually in principle, however, considering that material contents can get outdated, we may conservatively suppose its lifetime as 10 years without major revision. As for mini-projector, we may assume that such equipment gets depreciated in 5-year time. Presuming their equal use value in each

time-period, simply dividing the costs of materials and projector, their annual costs were 494USD and 346USD, respectively. Therefore, the total annual cost of HE was 16,462USD including the fixed cost. Per school annual running cost (excluding workshops) was 161USD or 183USD including the workshops and other fixed cost. Converting the annual cost to monthly cost, it was 13USD or 15USD, respectively. This monthly cost of our project was far less than that of the four weekly school hygiene education sessions and hygiene nudge construction project cited above[14] which was 127USD or 124USD per school, respectively, which would be 98USD or 96USD adjusting for the annual average inflation rate of 6.56% during 2013-2016 in Bangladesh.[50] The number of attended pupils in HE schools at the endline, thus assumed to have been treated during the project, were 15,622 (those in non-HE schools were 14,381). This makes the annual labour cost of para-teachers as well as of the total HE per attended student at 0.92USD and 1.05USD, respectively. As a reference, this was twice expensive than per student deworming cost of 0.49USD in large scale Tanzanian government project cited in [8].

In terms of cost-effectiveness of HE project, dividing the annual total HE cost per school (183USD) by the HE-treatment effect on *school hygiene practice* (0.36 as shown in Table 4 and Table 7) for instance, one percentage improvement in school hygiene was achieved by 5.08USD per school. Divided by the mean attended 173.8 students in an HE school at the endline, the respective cost-effectiveness was 0.029USD per pupil. Incorporating the effects of the composite *externality index* shown in Table 7, the total effect of HE became 0.63 points improvements in *school hygiene practice* per school, then one percentage improvement in school hygiene in HE schools was achievable by 2.90USD per school or 0.017USD per pupil, taking into account the spillover effects. The cost-effectiveness of one percentage improvement in school-level outcomes and in child-level outcomes are provided in Table 9 and Table 10, respectively. As a reference, the above-cited work [9] which conducted community intervention estimated the cost of health knowledge improvement by one percentage to be around 0.75~0.82USD, and personal hygiene improvement by one percentage to be

around 1.10~1.32USD per household. Although our case was per school/child, and the content of health education was not the same, comparing these HE costs, our school-based SBHE seem to be highly cost effective.

Table 9. Para-teacher and total HE cost in USD for 1% improvement in family-wise school health KAPB

	Cost per 1% improvement in school-level outcome	
	(S1) school hygiene practice & maintenance	
	para-teacher cost	total cost
HE effects	4.46	5.08
HE + spillover	2.55	2.90

Notes: Figures shown are calculated running (para-teachers) cost and total cost (including fixed cost) per school in USD per 1% improvement in outcomes for which estimated coefficient of *HE* and *eternality index* had statistical significance of at least 5% given in Table 7.

Table 10. Para-teacher and total HE cost in USD for 1% improvement in family-wise child health KAPB

	Cost per 1% improvement in child-level outcome					
	(P1) handwashing		(P2) dentalcare		(P3) overall hygiene	
	para-teacher cost	total cost	para-teacher cost	total cost	para-teacher cost	total cost
HE effects	0.04	0.05	0.05	0.06	0.04	0.05
HE + spillover	0.02	0.03	0.02	0.02	0.02	0.02
	(P6) knowledge		(P6E) knowledge + extra		(H2) cold-related symptoms	
	para-teacher cost	total cost	para-teacher cost	total cost	para-teacher cost	total cost
HE effects	0.02	0.02	0.05	0.05	-0.18	-0.21
HE + spillover	0.02	0.02	0.03	0.04	-0.09	-0.11

Notes: Figures shown are calculated running (para-teachers) cost and total cost (including fixed cost) per child in USD per 1% improvement in outcomes for which estimated coefficient of *HE* and *eternality index* had statistical significance of at least 5% given in Table 8.

Discussions and Conclusions

Our SBHE had positive effects on improved healthy/hygienic practice and behaviour both at the school and child levels. The evidence indicated possible establishment of new healthier norms. Cleaner and better maintained school infrastructure and self-hygiene, as sustained by the evidence, were expected to also serve as infection prevention from diseases and parasites. This was in part confirmed by children's health improvements related to cold symptoms. The fact that no impact was observed for

child anthropometry or child nutrition practice suggests that one-year SBHE alone was not sufficient in improving child nutrition intake. This resonates to the findings that neither educational nor food voucher provision was effective in nutrition practice on its own but only when provided together [51].The fact that the growth spurts happens at around puberty might also explain the no-effect on child anthropometry. Despite being a school-based intervention, our analysis showed that parents' literacy and particularly household wealth positively affected all child-level outcomes including child health and healthy behaviour as seen in [44]. While some outcomes such as handwashing habits were based on self-reports by the children, their knowledge regarding correct handwashing procedures performed and evaluated by the surveyors did support their responses. Also, the fact that children did not reply to have done the 'correct practice' in all demanded occasions buttressed the objectivity of their responses. The SBHE effects on dentalcare resonated with those of other successful SBHE studies [11,15,52]. Our results suggest that school-aged children were sufficiently amenable to new habit formation with possible peer effects in schools, achieved through a one-year intervention of weekly SBHE sessions with an emphasis on practical skill-building and context-relevant active learning.

Several plausible reasons limited our study findings particularly for the secondary outcomes of schooling and child health. Either a weekly SBHE for one year was insufficient to make improvement in these outcomes, insufficient statistical power to detect the effects given sample size of 180 schools and the actual effect size, or contamination due to spillover effects diluting the true effect measure. Other reasons could be the attrition of relatively unhealthy children that could have caused a downward bias for the estimated effects, or the cancelling-off effects between HE session and PE session, as HE session was carried out using PE class allotment. If PE classes, playing around, were assumed to have positive health effects, estimated HE effects could be biased toward zero. Concerning *schooling*, no treatment effect could be due to the likely undercounted dropout/repetition figures provided by the schools. The repetition figure provided by the school record amounting to 8.5% for grade 1~5 in the endline was far less than that of 20% in our randomly chosen surveyed pupils. Although the inclusion

of spillover/externality measures did not affect the effect size of other estimated coefficients, such spillover effects might have diluted the treatment effect estimate itself. Spillover effects were indeed statistically significant, and the closer and the more HE-schools, the larger was the effect of spillover. The period effect might have also captured part of the effects; as we have seen, there was a large general improvement from the baseline to the endline across most of the outcomes.

As for the strengths of our study, our SBHE was designed to overcome previously identified difficulties in conducting health education in terms of its cost, context relevance and effectiveness, and unmotivated teachers [9,33,53,54]. Our SBHE intervention employing para-teachers and using mobile projectors simultaneously overcame the problems of over-burdened primary school teachers, limited budgets, and knowledge gaps among the instructors. The use of mini-projectors with digital materials was also supportive for pupils to acquire healthy behaviours, by engaging them in the context relevant animations and images. The study rigorously conducted with randomly chosen samples and treatment allocation. Analysis-wise, we estimated the mean-standardised average treatment effect on the family of outcomes in order to avoid over identification of statistically significant results. Another strength and the novelty of our study was that we explicitly measured the spillover effects of SBHE. While spillover effects are generally considered as nuisance to the RCT evaluation, in actuality, they could disseminate part of SBHE without extra cost to other schools and pupils. Spillover effects could also reduce the de facto cost of intervention, as shown in our cost-effectiveness analysis. The results suggested that relatively uncostly SBHE could possibly induce healthy behaviours not only in treatment schools and their children, but also in other non-treated schools and pupils. Once acquired, healthy habits are non-costly financially and sustainable per se, and they can reinforce the positive impacts of other supply-based intervention, as proven by the better management and maintenance of school hygiene infrastructure. To the extent that Jhenaidah district was a typical rural area and that the samples were sufficiently representative of the population, the findings would be generalisable to other rural areas in Bangladesh. Although our study was for a limited period, further research looking into

how well such healthy environment and behaviour sustain would be warranted. Along with economic development, healthy behavioural changes and norms are expected to contribute to breaking the vicious circle of ill health, poor education and poverty, and to overcome a problem of low investment in own health toward sustainable health improvement.

References

1. Bangladesh Directorate General of Health Services. Health Bulletin 2010. Mohakhali Dhaka 1212: Government of the People's Republic of Bangladesh Ministry of Health and Family Welfare, Directorate General of Health Services; 2010.
2. Management Information System, Bangladesh Directorate General of Health Services. Health Bulletin 2020. Government of the People's Republic of Bangladesh Ministry of Health and Family Welfare; 2022. Available: https://old.dghs.gov.bd/images/docs/vpr/lhb_2020.pdf
3. Schools ill-equipped to provide healthy and inclusive learning environments for all children – UNICEF, WHO. 23 Jun 2022 [cited 27 Apr 2024]. Available: https://www.unicef.org/bangladesh/en/press-releases/schools-ill-equipped-provide-healthy-and-inclusive-learning-environments-all
4. Tomokawa S, Kasai T, Kobayashi T. Present conditions and problems of health education in Republic of Niger : The Importance of Health Education in Primary Schools in a developing country. In: Journal of International Development and Cooperation [Internet]. 31 Oct 2005 [cited 9 Aug 2020]. Available: http://ir.lib.hiroshima-u.ac.jp/00029772
5. Alam M, Winch PJ, Saxton RE, Nizame FA, Yeasmin F, Norman G, et al. Behaviour change intervention to improve shared toilet maintenance and cleanliness in urban slums of Dhaka: a cluster-randomised controlled trial. Trop Med Int Health. 2017;22: 1000–1011.
6. World Health Organization. Making every school a health-promoting school – Implementation Guidance. 22 Jun 2021 [cited 14 Mar 2023]. Available: https://www.who.int/publications-detail-redirect/9789240025073
7. Vaivada T, Sharma N, Das JK, Salam RA, Lassi ZS, Bhutta ZA. Interventions for Health and Well-Being in School-Aged Children and Adolescents: A Way Forward. Pediatrics. 2022;149: e2021053852M. doi:10.1542/peds.2021-053852M
8. Miguel E, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica. 2004;72: 159–217.
9. Mascie-Taylor CGN, Karim R, Karim E, Akhtar S, Ahmed T, Montanari RM. The cost-effectiveness of health education in improving knowledge and awareness about intestinal parasites in rural Bangladesh. Econ Hum Biol. 2003;1: 321–330. doi:10.1016/j.ehb.2003.08.001
10. Naqvi FA, Das JK, Salam RA, Raza SF, Lassi ZS, Bhutta ZA. Interventions for Neglected Tropical Diseases Among Children and Adolescents: A Meta-analysis. Pediatrics. 2022;149: e2021053852E. doi:10.1542/peds.2021-053852E
11. Akera P, Kennedy SE, Lingam R, Obwolo MJ, Schutte AE, Richmond R. Effectiveness of primary school-based interventions in improving oral health of children in low- and middle-income countries: a systematic review and meta-analysis. BMC Oral Health. 2022;22: 264. doi:10.1186/s12903-022-02291-2
12. Haleem A, Khan MK, Sufia S, Chaudhry S, Siddiqui MI, Khan AA. The role of repetition and reinforcement in school-based oral health education-a cluster randomized controlled trial. BMC Public Health. 2015;16: 2. doi:10.1186/s12889-015-2676-3

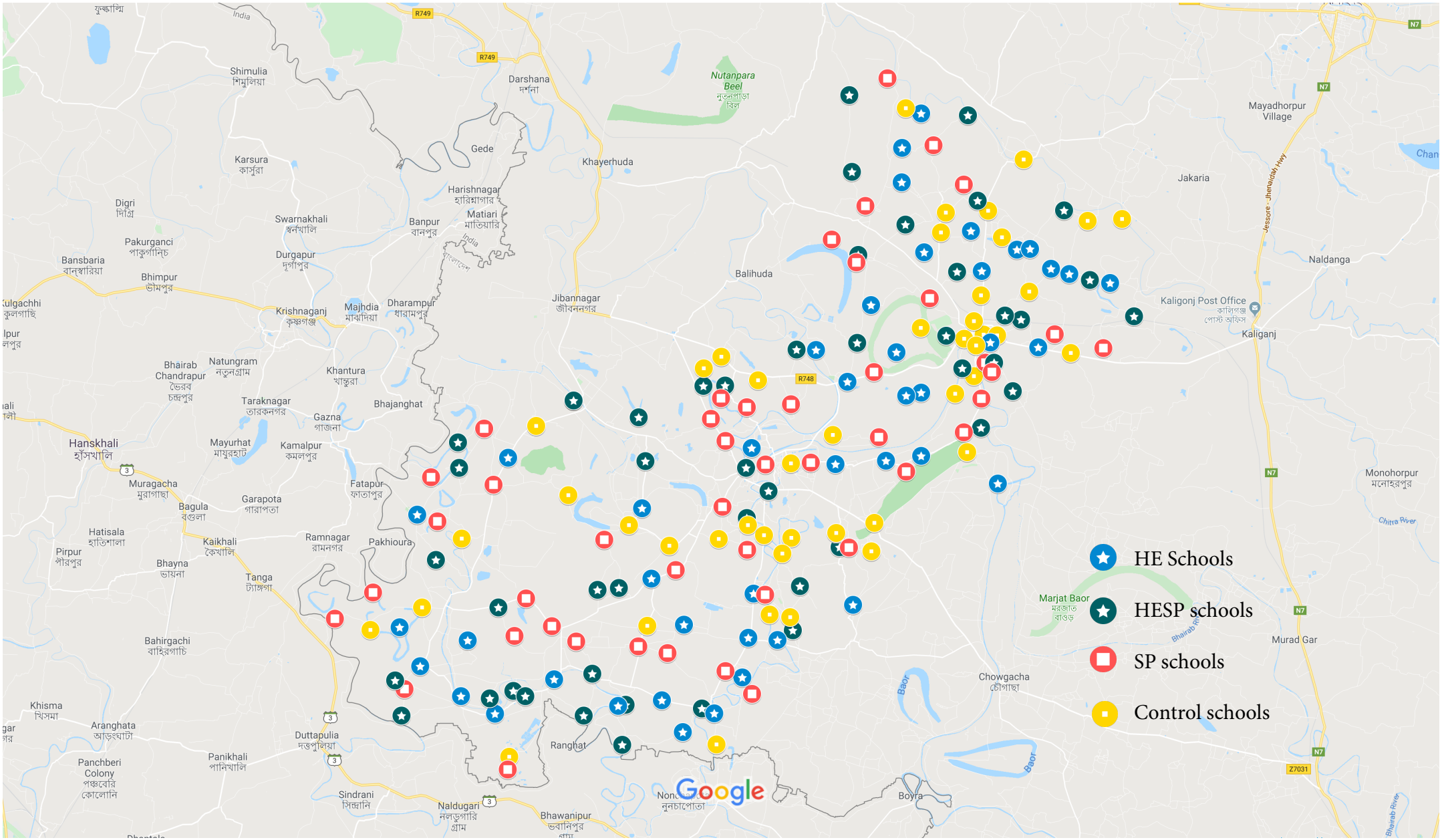
13. Luby SP, Agboatwalla M, Feikin DR, Painter J, Billhimer W, Altaf A, et al. Effect of handwashing on child health: a randomised controlled trial. Lancet Lond Engl. 2005;366: 225–233. doi:10.1016/S0140-6736(05)66912-7
14. Grover E, Hossain MK, Uddin S, Venkatesh M, Ram PK, Dreibelbis R. Comparing the behavioural impact of a nudge-based handwashing intervention to high-intensity hygiene education: a cluster-randomised trial in rural Bangladesh. Trop Med Int Health. 2018;23: 10–25. doi:10.1111/tmi.12999
15. Nyandindi U, Milén A, Palin-Palokas T, Robison V. Impact of oral health education on primary school children before and after teachers' training in Tanzania. Health Promot Int. 1996;11: 193–201. doi:10.1093/heapro/11.3.193
16. Kotb, M., Al-Teheawy, M., El-Setouhy, M., Hussein, H. Evaluation of a school-based health education model in schistosomiasis: A randomized community trial. East Mediterr Health J. 1998;4: 265–275.
17. Biswas D, Ahmed M, Roguski K, Ghosh PK, Parveen S, Nizame FA, et al. Effectiveness of a Behavior Change Intervention with Hand Sanitizer Use and Respiratory Hygiene in Reducing Laboratory-Confirmed Influenza among Schoolchildren in Bangladesh: A Cluster Randomized Controlled Trial. Am J Trop Med Hyg. 2019;101: 1446–1455. doi:10.4269/ajtmh.19-0376
18. Duflo E, Dupas P, Kremer M, Sinei S. Education And HIV/AIDS Prevention : Evidence From A Randomized Evaluation In Western Kenya. The World Bank; 2006. doi:10.1596/1813-9450-4024
19. Dupas P. Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya. Am Econ J Appl Econ. 2011;3: 1–34. doi:10.1257/app.3.1.1
20. Sato R, Takasaki Y. Peer Effects on Vaccination Behavior: Experimental Evidence from Rural Nigeria. Econ Dev Cult Change. 2019;68: 93–129. doi:10.1086/700570
21. Monsalve MN, Pemmaraju SV, Thomas GW, Herman T, Segre AM, Polgreen PM. Do Peer Effects Improve Hand Hygiene Adherence among Healthcare Workers? Infect Control Hosp Epidemiol. 2014;35: 1277–1285. doi:10.1086/678068
22. Baruah P, Boruah BB. Positive peer pressure and behavioral support. Indian J Posit Psychol. 2016;7: 241–243.
23. Ding W, Lehrer SF. Do Peers Affect Student Achievement in China's Secondary Schools? Rev Econ Stat. 2007;89: 300–312.
24. Lagomarsino BC, Gutman M, Freira L, Lanzalot ML, Lauletta M, Malchik LE, et al. Peer Pressure: Experimental Evidence from Restroom Behavior. Econ Inq. 2017;55: 1579–1584. doi:https://doi.org/10.1111/ecin.12437
25. Trogdon JG, Nonnemaker J, Pais J. Peer effects in adolescent overweight. J Health Econ. 2008;27: 1388–1399. doi:10.1016/j.jhealeco.2008.05.003
26. Fletcher JM. Peer Influences on Adolescent Alcohol Consumption: Evidence Using an Instrumental Variables/Fixed Effect Approach. J Popul Econ. 2012;25: 1265–1286.
27. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of 'habit-formation' and general practice. Br J Gen Pract. 2012;62: 664–666. doi:10.3399/bjgp12X659466
28. Royer H, Stehr M, Sydnor J. Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. Am Econ J Appl Econ. 2015;7: 51–84.
29. O'Donoghue T, Rabin M. Doing It Now or Later. Am Econ Rev. 1999;89: 103–124. doi:10.1257/aer.89.1.103
30. Verplanken B. Beyond frequency: Habit as mental construct. Br J Soc Psychol. 2006;45: 639–656. doi:10.1348/014466605X49122
31. Lally P, Jaarsveld CHM van, Potts HWW, Wardle J. How are habits formed: Modelling habit formation in the real world. Eur J Soc Psychol. 2010;40: 998–1009. doi:10.1002/ejsp.674
32. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340: c332. doi:10.1136/bmj.c332

33. Khandeparkar P, Pereira B, Sharma A, Gupta R, Shinde S, Patel V, et al. Promoting school climate and health outcomes with the SEHER multi-component secondary school intervention in Bihar, India: a cluster-randomised controlled trial. Lancet. 2018;392: 2465–2477.
34. Kling JR, Liebman JB. Experimental Analysis of Neighborhood Effects on Youth. SSRN Electron J. 2004 [cited 3 Sep 2022]. doi:10.2139/ssrn.600596
35. Clingingsmith D, Khwaja AI, Kremer M. Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering. Q J Econ. 2009;124: 1133–1170.
36. Duflo E, Glennerster R, Kremer M. Using Randomization in Development Economics Research: A Toolkit. In: Schultz T, Strauss J, editors. Amsterdam and New York: North Holland; 2008.
37. Glennerster R, Takavarasha K. Running Randomized Evaluations: A Practical Guide. Princeton and Oxford: Princeton University Press; 2013.
38. Angrist JD, Pischke J-S. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton and Oxford: Princeton University Press; 2009. Available: http://press.princeton.edu/titles/8769.html
39. Omura M, Venkatesh M, Khandaker I, Rahman A. Promoting Healthy Practices among Schools and Children in Rural Bangladesh: A Randomised Controlled Trial of Skill-Based Health Education. Manuscr Rev. 2024.
40. Bertrand M, Duflo E, Mullainathan S. How Much Should We Trust Differences-in-Differences Estimates? Q J Econ. 2004;119: 249–275.
41. Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA and London: MIT Press; 2010.
42. Cameron AC, Miller DL. A Practitioner's Guide to Cluster-Robust Inference. J Hum Resour. 2015;50: 317–372. doi:10.3368/jhr.50.2.317
43. McKenzie D. Beyond baseline and follow-up: The case for more T in experiments. J Dev Econ. 2012;99: 210–221. doi:10.1016/j.jdeveco.2012.01.002
44. Luby SP, Agboatwalla M, Hoekstra RM, Rahbar MH, Billhimer W, Keswick BH. DELAYED EFFECTIVENESS OF HOME-BASED INTERVENTIONS IN REDUCING CHILDHOOD DIARRHEA, KARACHI, PAKISTAN. Am J Trop Med Hyg. 2004;71: 420–427. doi:10.4269/ajtmh.2004.71.420
45. Madajewicz M, Pfaff A, van Geen A, Graziano J, Hussein I, Momotaj H, et al. Can information alone change behavior? Response to arsenic contamination of groundwater in Bangladesh. J Dev Econ. 2007;84: 731–754. doi:10.1016/j.jdeveco.2006.12.002
46. Jaime Torres MM, Carlsson F. Direct and spillover effects of a social information campaign on residential water-savings. J Environ Econ Manag. 2018;92: 222–243. doi:10.1016/j.jeem.2018.08.005
47. Fafchamps M, Vicente PC. Political violence and social networks: Experimental evidence from a Nigerian election. J Dev Econ. 2013;101: 27–48. doi:10.1016/j.jdeveco.2012.09.003
48. Duflo E, Saez E. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. Q J Econ. 2003;118: 815–841. doi:10.1162/00335530360698432
49. Varshney D, Joshi PK, Kumar A, Mishra AK, Kumar Dubey S. Examining the transfer of knowledge and training to smallholders in India: Direct and spillover effects of agricultural advisory services in an emerging economy. World Dev. 2022;160: 106067. doi:10.1016/j.worlddev.2022.106067
50. The World Bank. Inflation, consumer prices (annual %) - Bangladesh | Data. In: The World Bank Data [Internet]. 2021 [cited 26 Mar 2023]. Available: https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?end=2021&locations=BD&start=2001

51. Han Y, Kim HB, Park S. The Roles of Nutrition Education and Food Vouchers in Improving Child Nutrition: Evidence from a Field Experiment in Ethiopia. J Health Econ. 2021;80: 102545. doi:10.1016/j.jhealeco.2021.102545
52. Albandar JM, Buischi YA, Mayer MP, Axelsson P. Long-term effect of two preventive programs on the incidence of plaque and gingivitis in adolescents. J Periodontol. 1994;65: 605–610. doi:10.1902/jop.1994.65.6.605
53. Chaudhury N, Hammer J, Kremer M, Muralidharan K, Rogers FH. Missing in Action: Teacher and Health Worker Absence in Developing Countries. J Econ Perspect. 2006;20: 91–116. doi:10.1257/089533006776526058
54. Ramachandran V. Why School Teachers Are Demotivated and Disheartened. Econ Polit Wkly. 2005;40: 2141–2144.

Appendices

A-Figure 1. Project School Map in Jhenaidah, Bangladesh



Note: Map created using Google map.

A-Table 1. School-level analysis results: family-wise mean-standardised cross-cutting HESP treatment effect on five outcome families (N=360)

	primary outcomes		secondary outcomes			
	(S1) school hygiene practice & maintenance		(S2) hygiene infrastructure		(S3) schooling	
	β-coefficient [95%CI] [p-value]		β-coefficient [95%CI] [p-value]		β-coefficient [95%CI] [p-value]	
HE-treatment	0.19	0.19	0.09	0.1	0.13	0.13
	[-0.07,0.45]	[-0.07,0.45]	[-0.26,0.44]	[-0.24,0.44]	[-0.10,0.36]	[-0.10,0.36]
	[0.150]	[0.150]	[0.606]	[0.571]	[0.280]	[0.280]
HESP- treatment	0.37**	0.37**	-0.14	-0.15	0.01	0.01
	[0.10,0.64]	[0.10,0.64]	[-0.50,0.22]	[-0.50,0.21]	[-0.21,0.22]	[-0.21,0.22]
	[0.008]	[0.008]	[0.449]	[0.417]	[0.952]	[0.950]
SP-treatment	-0.16	-0.16	-0.02	-0.02	-0.04	-0.04
	[-0.42,0.09]	[-0.42,0.09]	[-0.38,0.35]	[-0.38,0.35]	[-0.29,0.20]	[-0.29,0.20]
	[0.211]	[0.212]	[0.932]	[0.933]	[0.727]	[0.727]
HE-group	0.02	0.02	-0.08	-0.06	-0.09	-0.09
	[-0.18,0.22]	[-0.18,0.22]	[-0.34,0.18]	[-0.31,0.20]	[-0.22,0.05]	[-0.22,0.05]
	[0.852]	[0.842]	[0.534]	[0.655]	[0.209]	[0.211]
HESP-group	-0.23	-0.23	0.16	0.15	0.09	0.09
	[-0.54,0.07]	[-0.54,0.07]	[-0.22,0.53]	[-0.22,0.52]	[-0.11,0.28]	[-0.11,0.28]
	[0.138]	[0.137]	[0.419]	[0.430]	[0.385]	[0.383]
SP-group	0.03	0.03	-0.05	-0.05	-0.1	-0.1
	[-0.18,0.24]	[-0.18,0.24]	[-0.33,0.23]	[-0.33,0.23]	[-0.26,0.05]	[-0.26,0.05]
	[0.759]	[0.758]	[0.718]	[0.727]	[0.193]	[0.192]
period	0.68***	0.68***	0.53***	0.52***	-0.03	-0.03
	[0.50,0.86]	[0.50,0.86]	[0.26,0.79]	[0.26,0.78]	[-0.21,0.15]	[-0.21,0.15]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.723]	[0.724]
school type	-0.16**	-0.15**	-0.04	0.02	0.06	0.05
	[-0.25,-0.06]	[-0.27,-0.04]	[-0.16,0.08]	[-0.12,0.16]	[-0.02,0.13]	[-0.03,0.14]
	[0.001]	[0.008]	[0.524]	[0.783]	[0.164]	[0.225]
female teacher ratio		0.01 [-0.18,0.21] [0.909]		0.21 [-0.05,0.47] [0.109]		-0.01 [-0.16,0.15] [0.939]
N	360	360	360	360	360	360

Note: Each column represents a separate regression on a family of outcomes estimated by seemingly unrelated regressions (SUR), estimated by a feasible generalised least squares (FGLS) estimator. β-coefficient is the mean standardised effect from SUR applying the difference-in-differences (DID) model, controlling for school type. An additional covariate is female teacher ratio. Outcome family compositions are: (S1) school hygiene practice: latrine cleaning rota, latrine cleaning days per week, classroom cleaning rota, classroom cleaning days per week, rubbish bin provision, latrine brush provision, rubbish disposal method, soap provision at handwashing facility; (S2) school hygiene infrastructure: clean classrooms, clean accessible latrines for boys; clean accessible latrines for girls, ratio of clean-usable latrine for boys, ratio of clean-usable latrine for girls, closeness of handwashing facility to the latrines (within 10~11 steps), easiness of obtaining water by younger children; (S3) schooling: absence rate for each grade1~5, repetition rate and dropout rate for each grade 2~5. Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence interval and p-value in brackets. 360 samples are 90 HE-schools and 90 control-schools in the baseline and endline.

A-Table 2. HE-treatment effects on each school level outcome (selected outcomes)

Outcome	HE-treatment effects				
	(1) pooled	(1e) endline	(2) DID	(3) logit	(4) ologit
rubbish bin in each classroom	0.34***	0.34***	0.43***	3.45**	3.32**
	[0.002]	[0.006]	[0.002]	[0.030]	[0.030]
rubbish disposal method	0.32**	0.32*	0.30*	2.55*	1.74
	[0.036]	[0.072]	[0.097]	[0.061]	[0.228]
soap for pupil at handwashing facility [(a)]	0.33***	0.33***	0.39***	3.10*	2.52**
	[0.002]	[0.001]	[0.006]	[0.054]	[0.035]
routine latrine cleaning	0.24***	0.24***	0.23**	3.60***	
	[0.000]	[0.000]	[0.011]	[0.008]	
latrine cleaning brush provision	0.11*	0.11***	0.13	5.29**	
	[0.062]	[0.009]	[0.124]	[0.022]	
weekly latrine cleaning days	0.53***	0.53**	0.49*		
	[0.008]	[0.022]	[0.074]		
SMC executed health related activity	0.12*	0.12*	0.09		
	[0.084]	[0.068]	[0.307]		
sealed latrine-pit	0.12	0.12	0.39**		
	[0.409]	[0.383]	[0.021]		
latrine detergent provided [(b)]		0.11**		2.44**	
		[0.039]		[0.042]	

Note: HE-treatment effects are given for each outcome shown in the first column for separate regression models (1) ~ (4). Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001, and p-value in brackets. Although results are not shown here, (a) for *soap for pupil at handwashing facility*, estimated coefficient of HE in HESP DID model is 0.400*, while that of HESP or SP is non-significant; (b) estimates for *latrine detergent provided* is available only for endline. Shown estimates for logit model and ordered-logit (ologit) are exponentiated (odds-ratio).

A-Table 3. Child-level analysis results: family-wise mean-standardised HE-treatment effect on nine outcome families for all children with additional covariates

	Primary Outcomes								
	(P1) handwashing			(P2) dentalcare			(P3) overall hygiene		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
HE-treatment	0.215***	0.211***	0.207***	0.169***	0.173***	0.168***	0.223***	0.225***	0.221***
	[0.13,0.30]	[0.12,0.30]	[0.12,0.30]	[0.08,0.26]	[0.08,0.27]	[0.07,0.26]	[0.14,0.31]	[0.14,0.31]	[0.14,0.30]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
HE-group	-0.053	-0.047	-0.05	-0.028	-0.025	-0.026	-0.060+	-0.056+	-0.057+
	[-0.12,0.02]	[-0.12,0.03]	[-0.12,0.03]	[-0.11,0.05]	[-0.10,0.05]	[-0.10,0.05]	[-0.12,0.00]	[-0.11,0.00]	[-0.12,0.00]
	[0.147]	[0.206]	[0.195]	[0.486]	[0.491]	[0.488]	[0.065]	[0.059]	[0.065]
period	0.603***	0.621***	0.641***	0.362***	0.375***	0.396***	0.491***	0.509***	0.529***
	[0.54,0.66]	[0.56,0.69]	[0.58,0.71]	[0.29,0.43]	[0.31,0.44]	[0.33,0.46]	[0.43,0.55]	[0.45,0.57]	[0.47,0.59]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
school type	-0.067*	-0.058*	-0.068*	-0.094**	-0.079**	-0.087**	-0.081**	-0.067**	-0.076**
	[-0.12,-0.01]	[-0.11,-0.01]	[-0.12,-0.01]	[-0.16,-0.03]	[-0.13,-0.02]	[-0.14,-0.03]	[-0.13,-0.03]	[-0.11,-0.02]	[-0.12,-0.03]
	[0.012]	[0.027]	[0.013]	[0.003]	[0.005]	[0.003]	[0.002]	[0.004]	[0.002]
sex	0.089***	0.097***	0.093***	0.040**	0.042**	0.036**	0.075***	0.077***	0.073***
	[0.07,0.11]	[0.08,0.12]	[0.07,0.11]	[0.01,0.07]	[0.02,0.07]	[0.01,0.06]	[0.06,0.09]	[0.06,0.09]	[0.06,0.09]
	[0.000]	[0.000]	[0.000]	[0.003]	[0.002]	[0.007]	[0.000]	[0.000]	[0.000]
child age	0.035***			0.036***			0.041***		
	[0.03,0.04]			[0.03,0.05]			[0.03,0.05]		
	[0.000]			[0.000]			[0.000]		
wealth index		0.108***			0.111***			0.104***	
		[0.09,0.12]			[0.09,0.13]			[0.09,0.12]	

		[0.000]			[0.000]			[0.000]	
parents' literacy			0.061***			0.078***			0.062***
			[0.05,0.08]			[0.06,0.10]			[0.05,0.08]
			[0.000]			[0.000]			[0.000]
N	16102	16171	16181	16104	16173	16183	16084	16153	16163
	(P4) clean hands			**(P4E) clean hands + ATP (endline)**			**(P5) nutrition**		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
HE-treatment	-0.004	-0.007	-0.008	0.06	0.07	0.07	-0.04	-0.042	-0.044
	[-0.11,0.10]	[-0.11,0.10]	[-0.11,0.10]	[-0.03,0.16]	[-0.02,0.17]	[-0.03,0.17]	[-0.09,0.01]	[-0.10,0.01]	[-0.10,0.01]
	[0.940]	[0.897]	[0.878]	[0.196]	[0.128]	[0.146]	[0.141]	[0.124]	[0.105]
HE-group	0.026	0.037	0.032				0.051*	0.057**	0.055*
	[-0.05,0.10]	[-0.04,0.11]	[-0.04,0.11]				[0.01,0.09]	[0.01,0.10]	[0.01,0.10]
	[0.501]	[0.332]	[0.400]				[0.020]	[0.008]	[0.010]
period	0.331***	0.374***	0.394***	0	0.03	0.01	-0.029	-0.023	-0.012
	[0.25,0.41]	[0.29,0.46]	[0.31,0.48]	[-0.10,0.10]	[-0.07,0.13]	[-0.08,0.11]	[-0.07,0.01]	[-0.06,0.02]	[-0.05,0.03]
	[0.000]	[0.000]	[0.000]	[0.971]	[0.545]	[0.806]	[0.146]	[0.233]	[0.554]
school type	-0.037	-0.024	-0.035				-0.01	-0.001	-0.007
	[-0.10,0.02]	[-0.08,0.03]	[-0.09,0.02]				[-0.04,0.02]	[-0.03,0.03]	[-0.03,0.02]
	[0.228]	[0.402]	[0.243]				[0.472]	[0.923]	[0.610]
sex	0.144***	0.149***	0.146***	0.07+	0.06	0.06	0.039***	0.041***	0.038***
	[0.11,0.17]	[0.12,0.18]	[0.12,0.17]	[-0.01,0.16]	[-0.02,0.14]	[-0.02,0.14]	[0.02,0.06]	[0.02,0.06]	[0.02,0.06]
	[0.000]	[0.000]	[0.000]	[0.070]	[0.163]	[0.150]	[0.000]	[0.000]	[0.000]
child age	0.066***			0.05***			0.019***		
	[0.06,0.07]			[0.03,0.08]			[0.01,0.02]		
	[0.000]			[0.000]			[0.000]		

wealth index		0.105***			0.07**			0.065***	
		[0.09,0.12]			[0.02,0.11]			[0.05,0.08]	
		[0.000]			[0.002]			[0.000]	
parents' literacy			0.052***			0.03			0.039***
			[0.03,0.07]			[-0.02,0.08]			[0.03,0.05]
			[0.000]			[0.185]			[0.000]
N	16099	16169	16178	867	867	867	16104	16173	16183

	(P6) knowledge			**(P6E) knowledge + extra (endline)**		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
HE-treatment	0.44***	0.44***	0.44***	0.19***	0.20***	0.20***
	[0.32,0.55]	[0.33,0.55]	[0.33,0.55]	[0.14,0.25]	[0.15,0.25]	[0.15,0.25]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
HE-group	-0.08	-0.08	-0.08			
	[-0.18,0.02]	[-0.17,0.02]	[-0.18,0.02]			
	[0.134]	[0.131]	[0.128]			
period	0.75***	0.79***	0.80***	-0.03	-0.02	-0.03
	[0.68,0.83]	[0.71,0.87]	[0.72,0.88]	[-0.09,0.02]	[-0.07,0.03]	[-0.08,0.02]
	[0.000]	[0.000]	[0.000]	[0.229]	[0.376]	[0.201]
school type	-0.02	-0.02	-0.02			
	[-0.08,0.04]	[-0.08,0.04]	[-0.08,0.04]			
	[0.516]	[0.600]	[0.506]			
sex	0.02+	0.03*	0.02*	0.02+	0.03*	0.03*
	[-0.00,0.04]	[0.00,0.05]	[0.00,0.05]	[-0.00,0.05]	[0.01,0.05]	[0.00,0.05]
	[0.053]	[0.021]	[0.040]	[0.059]	[0.017]	[0.024]
child age	0.06***			0.08***		

	[0.05,0.06]			[0.07,0.09]		
	[0.000]			[0.000]		
wealth index		0.06***			0.08***	
		[0.04,0.07]			[0.07,0.10]	
		[0.000]			[0.000]	
parents' literacy			0.04***			0.03***
			[0.02,0.06]			[0.01,0.04]
			[0.000]			[0.001]
N	16104	16173	16183	8991	8991	8991

	Secondary Outcomes								
	(H1) anthropometry			**(H2) cold-related symptoms**			**(H3) other illness**		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
HE-treatment	-0.019	-0.021	-0.021	-0.047+	-0.047+	-0.046+	0.02	0.017	0.018
	[-0.06,0.02]	[-0.06,0.02]	[-0.06,0.02]	[-0.10,0.01]	[-0.10,0.01]	[-0.10,0.01]	[-0.04,0.08]	[-0.05,0.08]	[-0.04,0.08]
	[0.313]	[0.282]	[0.285]	[0.082]	[0.085]	[0.088]	[0.528]	[0.587]	[0.568]
HE-group	0.011	0.014	0.011	0.031	0.029	0.029	-0.019	-0.018	-0.018
	[-0.05,0.08]	[-0.05,0.08]	[-0.06,0.08]	[-0.01,0.07]	[-0.01,0.07]	[-0.01,0.07]	[-0.08,0.04]	[-0.08,0.04]	[-0.08,0.04]
	[0.736]	[0.698]	[0.761]	[0.149]	[0.174]	[0.166]	[0.536]	[0.571]	[0.569]
period	0.192***	0.119***	0.140***	-0.031	-0.041+	-0.045*	-0.171***	-0.182***	-0.184***
	[0.16,0.22]	[0.09,0.15]	[0.11,0.17]	[-0.07,0.01]	[-0.08,0.00]	[-0.09,-0.00]	[-0.22,-0.12]	[-0.23,-0.14]	[-0.23,-0.14]
	[0.000]	[0.000]	[0.000]	[0.155]	[0.052]	[0.034]	[0.000]	[0.000]	[0.000]
school type	-0.008	0.008	-0.009	-0.004	-0.006	-0.005	-0.007	-0.009	-0.008
	[-0.06,0.05]	[-0.05,0.07]	[-0.07,0.05]	[-0.04,0.03]	[-0.04,0.02]	[-0.04,0.03]	[-0.05,0.03]	[-0.05,0.03]	[-0.05,0.03]
	[0.794]	[0.809]	[0.768]	[0.788]	[0.678]	[0.756]	[0.741]	[0.664]	[0.712]
sex	-0.046*	-0.050**	-0.048*	0.003	0.002	0.003	0.076***	0.075***	0.076***

	[-0.08,-0.01]	[-0.09,-0.01]	[-0.08,-0.01]	[-0.01,0.02]	[-0.02,0.02]	[-0.02,0.02]	[0.06,0.10]	[0.06,0.09]	[0.06,0.10]
	[0.010]	[0.007]	[0.011]	[0.721]	[0.853]	[0.761]	[0.000]	[0.000]	[0.000]
child age	-0.057***			-0.016***			-0.015***		
	[-0.07,-0.04]			[-0.02,-0.01]			[-0.02,-0.01]		
	[0.000]			[0.000]			[0.000]		
wealth index		0.092***			-0.021***			-0.015**	
		[0.07,0.11]			[-0.03,-0.01]			[-0.03,-0.00]	
		[0.000]			[0.000]			[0.005]	
parents' literacy			-0.030**			-0.014*			-0.006
			[-0.05,-0.01]			[-0.03,-0.00]			[-0.02,0.01]
			[0.008]			[0.014]			[0.346]
N	16062	16120	16130	16104	16173	16183	16094	16164	16173

Notes: Each column represents a separate regression on a family of outcomes applying seemingly unrelated regressions (SUR), estimated by a feasible generalised least squares (FGLS) estimator with cluster-robust standard errors (CRSE) except for (H2). β-coefficient is the mean-standardised effect from SUR applying the difference-in-differences (DID) model, controlling for school type, child sex and additional covariates, namely, *child age*, *wealth index*, and *parent literacy*. *Wealth index* is created through iterated principal factor, reflecting house structure materials, roof materials, number of rooms, latrine structure and materials, possession of electronic appliances, mobile phones and bikes. *Parents' literacy* reflects whether the child mother and father can read. For each outcome family, estimates are provided for all sample children and children present *in both* baseline and endline. Each indicator family includes the following variables: (P1) *handwashing practice*: handwashing habits (washing frequency in each occasion (before eating, after defecation, after playing); used substances (soap, ash, mud and/or water only); washing with soap in each occasion; wash with running water; correct washing procedure); (P2) *dental care practice*: dentalcare (frequency and use of brush/branch, toothpaste, etc.), type of materials used; (P3) *overall hygiene practice*: shoes/footwear wearing at school (frequency), shoes/footwear wearing at home (frequency in latrine and in courtyard), + P1 & P2; (P4) *clean hand*s: clean hands by observation, trimmed nails, clean nails; (P4E) *clean hand*s + *ATP*: additional hand cleanliness measured by ATP improvement rate (10% of samples); (P5) *nutrition practice*: breakfast habit, breakfast taken in 3 days, food taken in 3 days, ordered by the richness of nutrition score (none; carbohydrate and vitamins; carbohydrate, protein and plus); (P6) *health/hygiene knowledge*: handwashing procedure, breakfast significance; (P6E) *health/hygiene knowledge* + *extra:* additional knowledge measured only in the endline, i.e., putting water in latrine before defecating, oral rehydration solution (ORS) making, food pyramid; (I1) *anthropometry*: height-, weight-, BMI-z-score; (I2) *cold*: cold-related symptoms at present and in the past two-weeks, cough, breathing difficulty, sore throat, fever, running nose, congested nose; (I3) *other illness*: diarrhoea, stomachache, skin disease, fatigue, dizziness, appetite loss. Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence intervals and p-value in brackets.

A-Table 4. Child-level analysis results: family-wise mean-standardised cross-cutting HESP-treatment effect on selected outcome families for all children with additional covariates

	Primary Outcomes								
	(P1) handwashing			(P2) dentalcare			(P3) overall hygiene		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
HE-treatment	0.146*	0.142*	0.140*	0.174**	0.171**	0.168*	0.167**	0.164**	0.162*
	[0.03,0.27]	[0.02,0.26]	[0.02,0.26]	[0.04,0.31]	[0.04,0.30]	[0.04,0.30]	[0.04,0.29]	[0.04,0.29]	[0.04,0.29]
	[0.017]	[0.020]	[0.021]	[0.010]	[0.010]	[0.012]	[0.008]	[0.009]	[0.010]
HESP- treatment	0.203**	0.208***	0.199**	0.110+	0.119+	0.107	0.181**	0.186**	0.176**
	[0.08,0.32]	[0.09,0.33]	[0.08,0.32]	[-0.02,0.24]	[-0.01,0.25]	[-0.02,0.24]	[0.06,0.30]	[0.07,0.30]	[0.06,0.29]
	[0.001]	[0.001]	[0.001]	[0.097]	[0.070]	[0.108]	[0.003]	[0.002]	[0.004]
SP-treatment	-0.078	-0.079	-0.084	-0.041	-0.043	-0.047	-0.087	-0.089	-0.093
	[-0.20,0.04]	[-0.20,0.04]	[-0.20,0.03]	[-0.17,0.09]	[-0.17,0.09]	[-0.18,0.09]	[-0.20,0.03]	[-0.20,0.02]	[-0.21,0.02]
	[0.196]	[0.183]	[0.161]	[0.546]	[0.521]	[0.487]	[0.138]	[0.120]	[0.110]
HE-group	-0.04	-0.024	-0.031	-0.108+	-0.093+	-0.101+	-0.077	-0.062	-0.069
	[-0.14,0.06]	[-0.11,0.07]	[-0.13,0.06]	[-0.22,0.00]	[-0.19,0.01]	[-0.21,0.01]	[-0.17,0.02]	[-0.15,0.02]	[-0.16,0.02]
	[0.426]	[0.599]	[0.516]	[0.052]	[0.069]	[0.063]	[0.104]	[0.150]	[0.134]
HESP-group	-0.061	-0.052	-0.053	-0.039	-0.032	-0.032	-0.06	-0.051	-0.051
	[-0.17,0.04]	[-0.15,0.04]	[-0.15,0.05]	[-0.14,0.07]	[-0.13,0.07]	[-0.13,0.07]	[-0.15,0.03]	[-0.13,0.03]	[-0.14,0.04]
	[0.249]	[0.287]	[0.297]	[0.472]	[0.518]	[0.544]	[0.208]	[0.238]	[0.260]
SP-group	0.004	0.019	0.015	-0.095	-0.079	-0.083	-0.02	-0.004	-0.008
	[-0.10,0.11]	[-0.08,0.12]	[-0.09,0.12]	[-0.22,0.02]	[-0.19,0.03]	[-0.20,0.03]	[-0.12,0.08]	[-0.09,0.08]	[-0.10,0.09]
	[0.945]	[0.717]	[0.780]	[0.120]	[0.157]	[0.159]	[0.691]	[0.933]	[0.865]
period	0.637***	0.649***	0.671***	0.378***	0.394***	0.416***	0.528***	0.547***	0.567***
	[0.55,0.73]	[0.56,0.74]	[0.58,0.76]	[0.28,0.48]	[0.30,0.49]	[0.32,0.51]	[0.43,0.62]	[0.46,0.64]	[0.48,0.66]

	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
school type	-0.066*	-0.053*	-0.062*	-0.089**	-0.075**	-0.083**	-0.078**	-0.064**	-0.073**
	[-0.12,-0.01]	[-0.10,-0.01]	[-0.11,-0.01]	[-0.15,-0.03]	[-0.13,-0.02]	[-0.14,-0.03]	[-0.13,-0.03]	[-0.11,-0.02]	[-0.12,-0.03]
	[0.012]	[0.030]	[0.014]	[0.003]	[0.005]	[0.003]	[0.002]	[0.004]	[0.002]
sex	0.088***	0.090***	0.086***	0.039**	0.041**	0.036**	0.073***	0.076***	0.072***
	[0.07,0.11]	[0.07,0.11]	[0.07,0.10]	[0.01,0.06]	[0.02,0.07]	[0.01,0.06]	[0.06,0.09]	[0.06,0.09]	[0.05,0.09]
	[0.000]	[0.000]	[0.000]	[0.003]	[0.002]	[0.007]	[0.000]	[0.000]	[0.000]
child age	0.035***			0.037***			0.040***		
	[0.03,0.04]			[0.03,0.05]			[0.03,0.05]		
	[0.000]			[0.000]			[0.000]		
wealth index		0.101***			0.108***			0.100***	
		[0.09,0.12]			[0.09,0.12]			[0.09,0.11]	
		[0.000]			[0.000]			[0.000]	
parents' literacy			0.056***			0.075***			0.060***
			[0.04,0.07]			[0.06,0.09]			[0.05,0.07]
			[0.000]			[0.000]			[0.000]
N	16102	16171	16181	16104	16173	16183	16084	16153	16163

	(P6) knowledge			**(P6E) knowledge + extra (endline)**		
	β-coefficient [95%CI] [p-value]			β-coefficient [95%CI] [p-value]		
HE-treatment	0.395***	0.398***	0.396***	0.220***	0.220***	0.219***
	[0.24,0.55]	[0.25,0.55]	[0.24,0.55]	[0.14,0.30]	[0.14,0.30]	[0.14,0.30]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
HESP- treatment	0.401***	0.408***	0.402***	0.219***	0.224***	0.219***
	[0.25,0.56]	[0.25,0.56]	[0.25,0.56]	[0.14,0.30]	[0.14,0.31]	[0.14,0.30]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

SP-treatment	-0.07	-0.069	-0.071	-0.070+	-0.071+	-0.073+
	[-0.22,0.08]	[-0.22,0.08]	[-0.22,0.08]	[-0.15,0.01]	[-0.15,0.01]	[-0.15,0.00]
	[0.366]	[0.376]	[0.357]	[0.071]	[0.070]	[0.061]
HE-group	-0.051	-0.043	-0.047	-0.025	-0.014	-0.018
	[-0.19,0.09]	[-0.18,0.09]	[-0.19,0.09]	[-0.08,0.03]	[-0.07,0.04]	[-0.07,0.04]
	[0.479]	[0.541]	[0.510]	[0.402]	[0.614]	[0.514]
HESP-group	-0.085	-0.08	-0.08	-0.038	-0.03	-0.032
	[-0.22,0.05]	[-0.21,0.06]	[-0.22,0.06]	[-0.09,0.02]	[-0.08,0.02]	[-0.09,0.02]
	[0.230]	[0.247]	[0.251]	[0.192]	[0.263]	[0.250]
SP-group	0.016	0.028	0.026	0.001	0.014	0.011
	[-0.12,0.16]	[-0.11,0.17]	[-0.11,0.16]	[-0.06,0.06]	[-0.04,0.07]	[-0.04,0.07]
	[0.819]	[0.693]	[0.718]	[0.973]	[0.611]	[0.695]
period	0.781***	0.817***	0.829***	0.758***	0.795***	0.808***
	[0.68,0.89]	[0.71,0.92]	[0.72,0.93]	[0.70,0.81]	[0.74,0.85]	[0.75,0.86]
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
school type	-0.021	-0.016	-0.021	-0.034	-0.027	-0.033
	[-0.08,0.04]	[-0.08,0.04]	[-0.08,0.04]	[-0.08,0.01]	[-0.07,0.01]	[-0.07,0.01]
	[0.514]	[0.594]	[0.501]	[0.127]	[0.175]	[0.102]
sex	0.022+	0.026*	0.023*	0.021**	0.025***	0.023**
	[-0.00,0.04]	[0.00,0.05]	[0.00,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]
	[0.050]	[0.020]	[0.038]	[0.003]	[0.000]	[0.001]
child age	0.056***			0.055***		
	[0.05,0.06]			[0.05,0.06]		
	[0.000]			[0.000]		
wealth index		0.059***			0.066***	
		[0.04,0.07]			[0.05,0.08]	

		[0.000]			[0.000]	
parents' literacy			0.039***			0.028***
			[0.02,0.05]			[0.02,0.04]
			[0.000]			[0.000]
N	16104	16173	16183	16104	16173	16183

Notes: Each column represents a separate regression on a family of outcomes applying seemingly unrelated regressions (SUR), estimated by a feasible generalised least squares (FGLS) estimator with cluster-robust standard errors (CRSE) except for (H2). β-coefficient is the mean-standardised effect from SUR applying the difference-in-differences (DID) model, controlling for school type, child sex and additional covariates, namely, *child age*, *wealth index*, and *parent literacy*. *Wealth index* is created through iterated principal factor, reflecting house structure materials, roof materials, number of rooms, latrine structure and materials, possession of electronic appliances, mobile phones and bikes. *Parents' literacy* reflects whether the child mother and father can read. For each outcome family, estimates are provided for all sample children and children present *in both* baseline and endline. Selected outcome families are those for which treatment effect was found to be statistically significant in HE-treatment analysis given in S3 Table above; (P1) *handwashing practice*: handwashing habits (washing frequency in each occasion (before eating, after defecation, after playing); used substances (soap, ash, mud and/or water only); washing with soap in each occasion; wash with running water; correct washing procedure); (P2) *dental care practice*: dentalcare (frequency and use of brush/branch, toothpaste, etc.), type of materials used; (P3) *overall hygiene practice*: shoes/footwear wearing at school (frequency), shoes/footwear wearing at home (frequency in latrine and in courtyard), + P1 & P2; (P6) *health/hygiene knowledge*: handwashing procedure, breakfast significance; (P6E) *health/hygiene knowledge + extra:* additional knowledge measured only in the endline, i.e., putting water in latrine before defecating, oral rehydration solution (ORS) making, food pyramid. Significance level: + p<0.1, * p<0.05, ** p<0.01, ***p<0.001; 95% confidence intervals and p-value in brackets.

A-Table 5. HE-treatment effects on each child level outcome (selected outcomes

HE-Treatment Effects on Each Outcome at Individual-Level (Selected Outcomes)

	(1) Pooled		(1E) Endline		(2) FE		(3) DID		(4) ANCOVA		(5) Constrained Base		(6) Logit		(7) Ologit	
Outcome	all	*in both*	all	*in both*	all	*in both*	all	*in both*	all	*in both*	all	*in both*	all	*in both*	all	*in both*
weight z-score (net of clothes)	-0.020	*0.016*	-0.020	*0.016*	-0.033*	-0.033*	-0.032*	*-0.033**	-0.031	*-0.029*	-0.032**	*-0.030**				
	[0.648]	*[0.732]*	[0.648]	*[0.730]*	[0.094]	[0.094]	[0.094]	*[0.088]*	[0.208]	*[0.123]*	[0.047]	*[0.069]*				
cough	-0.014	*-0.016*	-0.0139	*-0.015*	-0.02	-0.02	-0.020	*-0.020*	-0.014	*-0.016*	-0.017*	*-0.017*	0.854	*0.847*		
	[0.200]	*[0.176]*	[0.197]	*[0.179]*	[0.313]	[0.313]	[0.191]	*[0.190]*	[0.178]	*[0.163]*	[0.088]	*[0.137]*	[0.216]	*[0.212]*		
breathing difficulty	-0.006	*-0.008**	-0.006	*-0.008**	-0.011*	-0.011*	-0.008*	*-0.011***	-0.006	*-0.007**	-0.006**	*-0.008***	0.631	*0.522**		
	[0.132]	*[0.078]*	[0.132]	*[0.078]*	[0.067]	[0.067]	[0.090]	*[0.040]*	[0.121]	*[0.083]*	[0.044]	*[0.020]*	[0.176]	*[0.095]*		
fever	-0.005	*-0.009*	-0.005	*-0.009*	-0.023**	-0.023**	-0.018**	*-0.023***	-0.006	*-0.01*	-0.01	*-0.014**	0.761	*0.698***		
	[0.467]	*[0.255]*	[0.466]	*[0.254]*	[0.043]	[0.043]	[0.043]	*[0.020]*	[0.4051]	*[0.1945]*	[0.1203]	*[0.0618]*	[0.108]	*[0.040]*		
cough in 2wks	-0.009	*-0.013*	-0.009	*-0.013*	-0.03	-0.03	-0.034	*-0.029*	-0.01	*-0.012*	-0.023*	*-0.021*	0.86	*0.87*		
	[0.593]	*[0.432]*	[0.591]	*[0.441]*	[0.269]	[0.269]	[0.132]	*[0.192]*	[0.563]	*[0.464]*	[0.070]	*[0.137]*	[0.212]	*[0.261]*		
breathing difficulty in 2wks	-0.007	*-0.008*	-0.007	*-0.008*	-0.014	-0.014	-0.012	*-0.013*	-0.007	*-0.008*	-0.009*	*-0.010**	0.68	*0.641*		
	[0.372]	*[0.330]*	[0.372]	*[0.331]*	[0.167]	[0.167]	[0.154]	*[0.136]*	[0.359]	*[0.323]*	[0.070]	*[0.072]*	[0.214]	*[0.177]*		
sore throat in 2wks	-0.005	*-0.005*	-0.005	*-0.005*	-0.011*	-0.011*	-0.010**	*-0.011***	-0.005	*-0.006*	-0.007**	*-0.008**	0.584*	*0.564**		
	[0.196]	*[0.222]*	[0.196]	*[0.222]*	[0.091]	[0.091]	[0.032]	*[0.033]*	[0.157]	*[0.1741]*	[0.047]	*[0.065]*	[0.094]	*[0.086]*		
running nose in 2wks	-0.005	*-0.005*	-0.005	*-0.005*	-0.037	-0.037	-0.041**	*-0.037**	-0.006	*-0.006*	-0.023*	*-0.02*	0.846	*0.86*		
	[0.749]	*[0.777]*	[0.749]	*[0.786]*	[0.174]	[0.174]	[0.050]	*[0.093]*	[0.662]	*[0.740]*	[0.070]	*[0.167]*	[0.118]	*[0.181]*		
diarrhea in 2wks	-0.007	*-0.021***	-0.007	*-0.021***	-0.021	-0.021	-0.0127	*-0.021**	-0.001	*-0.021***	-0.008	*-0.021***	0.899	*0.804**		
	[0.488]	*[0.032]*	[0.487]	*[0.032]*	[0.103]	[0.103]	[0.285]	*[0.082]*	[0.9071]	*[0.0313]*	[0.3032]	*[0.0207]*	[0.348]	*[0.064]*		
handwashing index	0.372**	*0.365***	0.372**	*0.364***	0.518**	0.518**	0.512**	*0.520***	0.392***	*0.371***	0.487***	*0.483****			1.309*	*1.299*
	[0.013]	*[0.015]*	[0.013]	*[0.016]*	[0.033]	[0.033]	[0.024]	*[0.022]*	[0.008]	*[0.012]*	[0.000]	*[0.000]*			[0.098]	*[0.102]*
handwashing index before eating*	0.195**	*0.198***	0.195**	*0.1980***	0.292**	0.292**	0.297***	*0.295***	0.210***	*0.207***	0.279***	*0.272****			1.448*	*1.414**
	[0.013]	*[0.015]*	[0.013]	*[0.016]*	[0.018]	[0.018]	[0.009]	*[0.011]*	[0.006]	*[0.011]*	[0.000]	*[0.000]*			[0.085]	*[0.092]*

handwashing index after defecation*	0.080**	*0.077***	0.080**	*0.077***	0.166**	0.166**	0.147**	*0.158****	0.085**	*0.076***	0.126***	*0.135****			1.391**	*1.420***
	[0.028]	*[0.035]*	[0.027]	*[0.036]*	[0.017]	[0.017]	[0.018]	*[0.009]*	[0.018]	*[0.040]*	[0.000]	*[0.000]*			[0.026]	*[0.012]*
handwashing index after playing*	0.097	*0.0886*	0.097	*0.088*	0.061	0.061	0.0687	*0.0689*	0.094	*0.086*	0.075*	*0.066*	1.068	*1.014*	1.027	*1*
	[0.278]	*[0.324]*	[0.278]	*[0.327]*	[0.626]	[0.626]	[0.552]	*[0.549]*	[0.292]	*[0.335]*	[0.086]	*[0.183]*	[0.799]	*[0.957]*	[0.878]	*[0.999]*
handwashing with soap before eating	0.067**	*0.068***	0.067**	*0.068***	0.114***	0.114***	0.114***	*0.113****	0.070***	*0.069***	0.103***	*0.101****	1.879***	*1.860****		
	[0.012]	*[0.015]*	[0.012]	*[0.015]*	[0.001]	[0.001]	[0.000]	*[0.001]*	[0.010]	*[0.016]*	[0.000]	*[0.000]*	[0.001]	*[0.002]*		
handwashing with soap after defecation	0.032***	*0.033****	0.032***	*0.033****	0.064**	0.064**	0.056**	*0.063****	0.033***	*0.033****	0.047***	*0.053****	1.560***	*1.711****		
	[0.008]	*[0.007]*	[0.008]	*[0.007]*	[0.018]	[0.018]	[0.016]	*[0.008]*	[0.0057]	*[0.008]*	[0.000]	*[0.000]*	[0.006]	*[0.003]*		
handwashing with soap after playing	0.025	*0.034*	0.025	*0.034*	0.062**	0.062**	0.0480**	*0.061***	0.031	*0.038**	0.042***	*0.053****	1.554**	*1.767***		
	[0.251]	*[0.139]*	[0.250]	*[0.140]*	[0.016]	[0.016]	[0.038]	*[0.012]*	[0.160]	*[0.095]*	[0.000]	*[0.000]*	[0.041]	*[0.010]*		
handwashing with running water	0.012	*0.005*	0.012	*0.0054*	0.023	0.023	0.025*	*0.024*	0.017	*0.011*	0.021***	*0.017***	2.559	*9.978***	1.455*	*1.472*
	[0.414]	*[0.703]*	[0.413]	*[0.695]*	[0.154]	[0.154]	[0.091]	*[0.105]*	[0.2404]	*[0.4374]*	[0.0052]	*[0.0395]*	[0.140]	*[0.040]*	[0.084]	*[0.122]*
handwashing procedures	0.774***	*0.821****	0.774***	*0.820****	0.850***	0.850***	0.807***	*0.838****	0.782***	*0.815****	0.801***	*0.843****	2.758***	*2.933****	3.532***	*3.796****
	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.000]	[0.000]	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.003]	*[0.004]*	[0.000]	*[0.000]*
tooth brushing frequency	0.071***	*0.074****	0.071***	*0.074****	0.094***	0.094***	0.091***	*0.0236*	0.069***	*0.065***	0.087***	*0.087****			2.007***	*1.992****
	[0.006]	*[0.008]*	[0.006]	*[0.008]*	[0.005]	[0.005]	[0.002]	*[0.105]*	[0.007]	*[0.018]*	[0.000]	*[0.000]*			[0.007]	*[0.008]*
dental-care index	0.124***	*0.113****	0.124***	*0.113****	0.148***	0.148***	0.151***	*0.087****	0.131***	*0.115****	0.143***	*0.134****	1.524***	*1.493***	1.717***	*1.713****
	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.001]	[0.001]	[0.000]	*[0.005]*	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.002]	*[0.013]*	[0.000]	*[0.000]*
wearing footwear at school	-0.001	*0.001*	-0.001	*0.001*	0.033*	0.033*	0.0238	*0.031***	0.003	*0.005*	0.017**	*0.024****	1.2	*1.367*		
	[0.898]	*[0.843]*	[0.897]	*[0.861]*	[0.075]	[0.075]	[0.115]	*[0.043]*	[0.6685]	*[0.3400]*	[0.0226]	*[0.0046]*	[0.409]	*[0.184]*		
wearing footwear at home	0.105***	*0.085***	0.105***	*0.084***	0.164**	0.164**	0.155***	*0.155****	0.105***	*0.075***	0.137***	*0.129****	3.032***	*2.780***	1.324**	*1.352***
	[0.005]	*[0.020]*	[0.005]	*[0.020]*	[0.014]	[0.014]	[0.007]	*[0.006]*	[0.005]	*[0.038]*	[0.000]	*[0.000]*	[0.010]	*[0.042]*	[0.035]	*[0.025]*
water latrine before defecate [(a)]	0.069*	*0.080**	0.069*	*0.080**							0.069*	*0.07*	1.13	*1.189**	1.168*	*1.196**
	[0.087]	*[0.055]*	[0.087]	*[0.055]*							[0.089]	*[0.100]*	[0.166]	*[0.073]*	[0.085]	*[0.055]*
ORS making knowledge[(a)]	0.238***	*0.259****	0.238***	*0.259****							0.238***	*0.252****	1.726***	*1.740****	1.713***	*1.728****
	[0.000]	*[0.000]*	[0.000]	*[0.000]*							[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.000]	*[0.000]*

breakfast significant	-0.006	*-0.004*	-0.006	*-0.0042*	0.037	0.037	0.043*	*0.035*	0.135***	*-0.004*	0.026***	*0.023***	0.891	*0.813*		
	[0.173]	*[0.189]*	[0.171]	*[0.182]*	[0.172]	[0.172]	[0.079]	*[0.136]*	[0.000]	*[0.1928]*	[0.0060]	*[0.0450]*	[0.648]	*[0.506]*		
breakfast eaten today	-0.0012	*0.003*	-0.0012	*0.0027*	-0.026	-0.026	-0.0302	*-0.0251*	0.098***	*0.004*	-0.01	*-0.005*	0.523**	*0.442****	0.906	*0.929*
	[0.937]	*[0.869]*	[0.937]	*[0.870]*	[0.615]	[0.615]	[0.115]	*[0.229]*	[0.0000]	*[0.8328]*	[0.4249]	*[0.7170]*	[0.028]	*[0.009]*	[0.256]	*[0.432]*
cleaned school latrine	0.095***	*0.109****	0.095***	*0.109****	0.194	0.194	0.104***	*0.111****	-0.001	*0.108****	0.101***	*0.113****	1.725***	*1.743****		
	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.307]	[0.307]	[0.000]	*[0.000]*	[0.930]	*[0.000]*	[0.000]	*[0.000]*	[0.000]	*[0.000]*		
seen other pupils clean school latrine	0.136***	*0.118****	0.136***	*0.117****	0.109***	0.109***	0.131***	*0.222****	0.003	*0.116****	0.132***	*0.111****	1.833***	*1.685****		
	[0.000]	*[0.000]*	[0.000]	*[0.000]*	[0.002]	[0.002]	[0.000]	*[0.001]*	[0.6103]	*[0.000]*	[0.000]	*[0.000]*	[0.000]	*[0.001]*		
swimming frequency	-0.059	*-0.042*	-0.059	*-0.042*	-0.024	-0.024	-0.039	*-0.026*	-0.046	*-0.032*	-0.043*	*-0.026*	0.937	*0.99*	0.912	*0.958*
	[0.271]	*[0.451]*	[0.271]	*[0.445]*	[0.606]	[0.606]	[0.363]	*[0.574]*	[0.278]	*[0.474]*	[0.073]	*[0.331]*	[0.563]	*[0.932]*	[0.351]	*[0.680]*

Note: HE-treatment effects are given for each outcome shown in the first column for separate regression models (1) ~ (6), with significance level * p<0.10, ** p<0.05, *** p<0.01, and p-value in brackets. Models (1) pooled OLS, (1E) endline OLS, (2) child-level fixed effects (FE); (3) DID estimation with multilevel random-effects (RE) , (6) logit and (7) ordered-logit (ologit) apply cluster-robust standard-error (CRSE).; models (4) Analysis of Covariance (ANCOVA) and (5) Constrained Baseline Analysis (CBA) control for intracluster correlation by adjusting for baseline cluster-mean, and (4) *in both* and (5) additionally adjust for individual-mean (for variable available only in endline, only cluster-mean adjustment is applied). Estimates for logit model and ordered-logit (ologit) are estimated in DID form and results shown are exponentiated (odds-ratio) ; logit results are also shown for ordered variables for valid estimates with sufficient 0 observations. The number of observations shown is general number and may vary by specific outcome variables. (a) *ORS making knowledge* is available only in endline. For ANCOVA *all* estimation, the presented results are adjusted for baseline cluster mean as well as individual baseline value.

Model Specification: Estimations are conducted for (1) pooled and (1e) endline only data by OLS estimator, (2) child-level FE and (3) DID estimation with child-level RE using maximum-likelihood (ML) estimator. For binary or ordered response variables, (6) logit and (7) ordered-logit (ologit) estimations by maximum likelihood estimator are also applied. While models (1)~(3), (5) and (6) apply CRSE, as pointed out by Angrist and Pischke (2009), Cameron and Miller (2015), and others, CRSE is not a panacea. McKenzie (2012) points to the fact that with a single baseline and follow-up, it would require twice the sample size in DID to get the same power as obtained with ANCOVA. The benefits and procedures of applying ANCOVA as well as baseline-constrained model are highlighted in a simulation by Hooper et al. (2018). As detailed in McKenzie (2012), instead of DID which attempts to fully control for the non-random differences in baseline that will likely jeopardize predictive power, we apply Analysis of Covariance (ANCOVA) estimation which is reported to be more efficient that DID or post estimator which uses only the endline data (Frison and Pocock 1992).

$(1A)\ Y_{ijt} = \delta \cdot t + \beta \cdot T_{ijt} + \varphi \cdot \bar{Y}_{j_{t=0}} + \upsilon \cdot Y_{i_{t=0}} + \varepsilon_{ijt},$

This method can adjust for the baseline cluster mean $\bar{Y}_{j_{t=0}}$ or baseline individual value of the outcome variable$Y_{i_{t=0}}$, or both baseline cluster mean and baseline individual value for more precise results (Klar and Darlington 2004; McKenzie 2012; Hooper et al. 2018). As an alternative to ANCOVA, Constrained Baseline Analysis (CBA) is also applied which allows for random effects of cluster as well as individual nested within cluster. CBA can be applied to all pupils at the endline regardless of having their data taken at the baseline (Hooper et al. 2018). In order to discern possible effects of attrition and replacement, as well as additional class 1 inclusion, we also present the results of estimation over children of closed cohort who were present in both the baseline and endline surveys (*in-both*).