DUBAI CARES WASH IN SCHOOLS INITIATIVE IN MALI

Impact Evaluation Report

January 2013 – May 2014

EMORY UNIVERSITY | Center for Global Safe Water

Prepared by:

Victoria Trinies, MPH

Sabah Ghulamali

Matthew C. Freeman, MPH, PhD

Department of Environmental Health
Rollins School of Public Health
Emory University
1518 Clifton Rd NE,
Atlanta, Georgia USA 30322

With Support from
Jérémie Toubkiss and Heather Murphy.

We would like to thank the UNICEF, WaterAid, CARE, Oxfam, and Save the Children teams for their support, specifically Jérémie Toubkiss, Yagouba Diallo, Seydou Niafo, Toureba Keita, Assitan Sogoré, Salimata Togola, Fatoumata Haidara, Mamadou Diallo, Zoumana Cisse, Ousmane Haidara, Thierno Bocoum, Alexandra Blason and Sue Cavill. Thanks to Sarah Porter who supported the initial proposal development. Thanks as well to the tireless research team and the Government and people of Mali.

July 2015

TABLE OF CONTENTS

List of acronyms

CAP	Local-level education district (*Centre d'animation pédagogique*)
CLTS	Community-led total sanitation
CNRST	*Centre National de la Recherche Scientifique et Technique*
DCIM	Dubai Cares WASH in Schools Initiative in Mali
IEC	Information, education, communication
IRB	Institutional Review Board
NGO	Non-governmental organization
RCT	Randomized controlled trial
SMC	School Management Committee (*Comité de gestion scolaire)*
UNICEF	United Nations Children's Fund
WASH	Water, sanitation, and hygiene

EXECUTIVE SUMMARY

This report presents key output, outcome, and impact findings from an evaluation of the Dubai Cares Water, Sanitation and Hygiene (WASH) in Schools Initiative in Mali (DCIM). This program, funded by the philanthropic foundation Dubai Cares and implemented by CARE Mali, Oxfam GB, Save the Children US, UNICEF-Mali, and WaterAid Mali, provided a comprehensive suite of school WASH infrastructure, supplies, behavior change and training activities to 916 primary and secondary schools across Mali with outreach into the surrounding communities.

We conducted a matched control evaluation to assess the impact of the DCIM program on pupil absence and health. Data were collected from 100 beneficiary schools participating in the DCIM program and 100 comparison schools in the same educational districts with similar WASH conditions and school population size. Schools were visited every six to eight weeks for a total of five to six visits. At each school visit, enumerators conducted interviews with 40 pupils in grades 3-6 and conducted a roll call of all pupils in grades 1-6. Enumerators also conducted interviews with school directors and performed structured observations of WASH facilities and supplies, as well as observation of pupil handwashing practices.

This evaluation provided evidence that the DCIM intervention improved the WASH environment in beneficiary schools, and that those improvements were sustained over two years after implementation of the program. We found that the program produced changes in pupils' handwashing and open defecation practices. Our data suggest that the program had a positive effect on pupils' self-reported incidence of diarrhea and respiratory infection symptoms. However, we did not see evidence of an effect on self-reported absence in the past week, and pupils in beneficiary school were more likely than pupils in comparison schools to be absent according to roll call data.

We found evidence that the DCIM program had a positive impact on pupil health. Pupil in the beneficiary group were 29% less likely to report having had diarrhea in the past week compared to pupils in comparison schools (Odds Ratio [OR] 0.71, 95% Confidence Interval [CI] 0.60-0.85), and were 25% less likely to report having had symptoms of respiratory infections (OR 0.75, 95%CI 0.65-0.86). Pupils in beneficiary schools were also 27% less likely to report diarrhea as cause for absence in the past week than pupils in comparison schools (OR 0.73, 95%CI 0.56-0.94).

We did not find evidence that the DCIM program had a positive impact on overall pupil absenteeism. There was no significant difference in the rates of self-reported absence among pupils in beneficiary and comparison schools (OR 0.93, 95%CI 0.79-1.09). Although we did see lower roll call absenteeism in the beneficiary schools after the intervention compared to baseline, pupils in beneficiary schools were 23% more likely to be absent at roll call than were pupils in comparison schools (OR 1.23, 95%CI 1.06-1.42). Although the increased rate of roll call absenteeism in beneficiary schools may have been related to the intervention, it is more likely that this is the result of an imbalance in unobserved confounders, or preferential enrollment of marginalized pupils in schools with greater WASH access. These imbalances at baseline can be attributed to the matched control study design, which unlike a randomized control design that randomly allocates schools to either intervention or control, relies on observed characteristics to assign schools to the control group. The design was a major limitation of the study.

We assessed how well schools performed at meeting four key WASH standards and fifteen related criteria at different time points following implementation of the DCIM program. These standards and associated criteria were established by the DCIM partners and included water access, sanitation, handwashing, and the presence of basic hygiene supplies. DCIM beneficiary schools performed better across all four of these compared to both the situation of the beneficiary schools at baseline and the group of comparison schools.

Beneficiary schools met the water access standard at 81.8% of evaluation visits, which required the presence of an improved and functional water point on the school grounds. There did not appear to be any decrease in the percentage of schools with functional water point as time since implementation increased.

Beneficiary schools also performed well at sustaining the presence of basic hygiene kit supplies and were observed to have handwashing soap, detergent or bleach, at least one anal cleansing kettle, and at least one drinking water container at 70.4% of evaluation visits. Beneficiary schools were somewhat more likely to have durable supplies, such as water containers, anal cleansing kettles, trash bins, drinking cups, and brooms (present at 94-98% of observations), compared to consumable supplies such as soap, detergent, and bleach (present at 71-82% or observations).

Beneficiary schools were less likely to meet the handwashing standard than water and hygiene kit related standards. The handwashing standard required the presence of at least one handwashing container with water and soap. It was met at 57.9% of all observations during the evaluation period. The greatest barriers to meeting this standard were ensuring that the handwashing containers contained water and that soap was made available to pupils. There was evidence that soap was somewhat less likely to be available and handwashing containers were less likely to be filled as time since implementation increased.

The schools were least likely to meet the sanitation standard, which required that there be at least one improved latrine per 70 pupils that was clean (no evidence of feces or pools of stagnant water or urine and the pit not full), safe (no large crack or vibrations in the floors, walls or roof), and sex-separated in practice. Beneficiary schools met this standard at 47.0% of observations. Schools were less likely to meet the safety and cleanliness criteria and to have sex-separated latrines as time since implementation increased.

Pupil behavior change also improved as a result of the DCIM program. Open defecation was observed in 23.8% of beneficiary schools on average, a decrease from 61.9% at baseline, and in 75.3% of comparison schools on average. The overall percentage of schools where no open defecation was observed appeared to be constant over time, with the lowest average rates of open defecation seen among beneficiary schools in the period 24+ month after implementation. Rates of observed handwashing with soap when soap was available also improved; pupils in beneficiary schools were observed to wash their hands with soap 57.7% of the time on average, compared to 39.6% at baseline. The practice of handwashing with soap appeared to be sustained over time.

We assessed several indicators related to governance, and beneficiary schools performed better across all governance indicators than did the schools at baseline and the comparison schools. Beneficiary schools met the criteria for an active and participatory school management committee (SMC) at 43.5% of observation visits. On average 36.3% of beneficiary schools reported having a budget dedicated to WASH, and beneficiary schools reported having spent money on WASH at 65.4% of observations. A teacher, pupil or parent was designated with responsibility for WASH at 92.3% of beneficiary schools on average. WASH lessons had been taught in all classes in 61.0% of beneficiary schools.

A. BACKGROUND & INTRODUCTION

In 2010, a consortium of partners including CARE Mali, Oxfam GB, Save the Children US, UNICEF-Mali, and WaterAid Mali, with financial support from the philanthropic foundation Dubai Cares, launched the Dubai Cares WASH in Schools Initiative in Mali (DCIM) to implement water, sanitation and hygiene (WASH) programming in schools across the country. The program was implemented in coordination with the Ministry of Primary Education and in conjunction with the Ministries of Sanitation, Water, and Health.

The DCIM partners identified public and community primary schools for inclusion in the program in collaboration with the Ministry of Education. Priority was given to schools with larger enrollment in order to maximize the coverage of the program. Schools were selected from six of Mali's nine regions: Bamako district, Gao, Koulikoro, Mopti, Sikasso, and Timbuktu. Location of schools within each region was targeted to areas where the partners already worked. Ultimately 916 schools participated in the program, including 734 primary schools and 182 secondary schools, and the program covered more than 25% of school going children in the six regions.

The DCIM partners implemented their programs under common standards, guidelines, and implementation targets but developed partner-specific implementation strategies. Core activities of the project at the school level ("intervention package") included:

- Installing or rehabilitating improved water points on the school compound.
- Installing or rehabilitating latrines so that there was at least one improved latrine for every 70 boys and girls on average.
- Distributing hygiene kits, which included handwashing and drinking water containers, soap, anal cleansing kettles, trash bin, and cleaning supplies such as brooms and disinfectant.
- Information, education, communication (IEC) activities and training, including:
    - Promoting good WASH practices and behavior change at school and within the community;
    - Providing training to teachers and school management committees;
    - Setting up school hygiene clubs or children's government; and
    - Establishing management systems to ensure sustained financing, monitoring and maintenance.

Differences in implementation included strategies for behavior change communication, inclusion of surrounding communities in behavior change activities, types of handwashing systems and of drinking water containers in classrooms, design of water points and latrines, construction of urinals, models for organizing responsibility and oversight of construction, and added components such as schools fencing, deworming, nutritional supplementation, and girls' clubs.

Implementation in schools overseen by CARE, Oxfam and WaterAid started in January 2011 and continued until June 2014. The program was implemented by UNICEF and

Save the Children in phases over this period. Implementation strategies were revised over time based on lessons learned.

In January 2012, armed conflict broke out in the northern part of the country. Intervention activities in Gao and Timbuktu regions ceased and surveillance was not feasible; new schools were selected in southern regions to replace them.

The Dubai Cares Foundation contracted researchers from Emory University to design and execute an independent evaluation of the impact of the program on pupil health and academic performance, as well as provide outcome monitoring data. This report presents findings from that evaluation, which was carried out from January 2013 through May 2014.

1. EVALUATION OBJECTIVES

The purpose of this evaluation was to quantify the impact of the DCIM program on the educational achievement and health status of students in beneficiary schools. As a secondary objective, program output and outcome data were also collected in order to determine the success and sustainability of the program at achieving implementation targets and promoting behavior change.

Few studies have assessed the impacts of school WASH on health and education attainment using rigorous methodology. Of the few existing studies, none have attempted to quantify impacts at the scale of the country-wide DCIM program. A rigorous evaluation of the health and educational impacts of the program will therefore provide a valuable contribution to the global knowledge base of school WASH impacts and contribute to global advocacy efforts for WASH in schools.

2. DATA COLLECTION METHODS

BASELINE

In March and April 2011, before the start of program activities, the five DCIM partners conducted a baseline survey among all schools that had been selected to participate in the program. At each school, enumerators carried out structured observations of school WASH facilities, structured observation of handwashing practices, interview with the school director, and review of school attendance and academic records.

We used simulation-based estimates to assess power for our primary outcome of interest [1]. Our study was powered to detect a significant change of 20% in the odds of roll call absence using at least 4 follow-up visits. Little data on absence was available in Mali, but we estimated a pre-intervention prevalence of 13% using pilot data. We calculated this anticipated detectible difference, given a sample size of 90 clusters per arm, which was determined by the available budget and assumed 80% power, $\alpha = 0.05$, and an estimate of 250 pupils per school. Given the limited data on absence in Mali and a lack of information on pupil variance in absence, we reconfirmed our power following collection of baseline data using simulation based estimation to confirm our sample size and establish the required number of follow-up rounds, applying the within pupil intra-cluster correlation (ICC) of 0.262 and a within school ICC of 0.065. While four follow-up visits were deemed sufficient, given the seasonality of absence and potential delays in observed benefits, we planned for at least five follow-up rounds.

In a subsample of 90 beneficiary schools, up to 80 pupils were selected to provide additional data on pupil health and absenteeism. In order to ensure that the sample reflected the overall makeup of program schools, all beneficiary schools were stratified by partner, region, and school size. Schools were then randomly selected for inclusion in the baseline health and absenteeism sample. Pupils in selected schools were asked to provide a recall of absence in the previous seven days and diarrhea (defined using local terminology) in the previous two days. Attendance data were also collected school-wide in this subsample through roll-call and review of daily school attendance reports. In schools where there were handwashing facilities, pupil handwashing practices after using the latrines were observed during the morning school break.

The five partners also collected baseline data from 92 comparison schools that had not been selected for inclusion in the DCIM program and which had not benefitted from any WASH programs in the past five years. These schools were matched to beneficiary schools based on their location in the same educational districts, known as *centres d'animation pédagogiques* (CAP), as the beneficiary schools and school size.

EVALUATION

We employed a cluster-matched longitudinal design for the impact evaluation. The primary outcome of the evaluation was absence measured using pupil roll call data collected by the study team.

SAMPLE SIZE AND SAMPLING

We powered our study based on the ability to detect a significant change in one-day roll call absence (primary outcome), self-reported absence in the past seven days, and diarrhea recall in the past seven days. The mean risk at baseline and coefficient of variation was estimated using baseline data.

The original intention of the impact evaluation was to use the sub-sample of schools that had provided additional pupil information during the 2011 baseline. However, the conflict in the north made 32 of these schools inaccessible and lost to follow-up, and 16 of the remaining schools did not meet the criteria for a complete intervention package. Ultimately 42 of the beneficiary schools and 9 control schools that participated in the baseline were retained to participate in the impact evaluation.

Beneficiary schools with the following criteria were eligible for inclusion in the impact evaluation:

- Public or community primary schools
- Had received or were scheduled to receive a complete DCIM intervention package by June 2013. A "complete package" includes:
    - Hardware: construction or rehabilitation of a functional and improved water point on the school compound and one improved latrine per 70 boys and 70 girls; distribution of a hygiene kits consisting of soap, brooms, handwashing containers, drinking water container, disinfectant, and water kettles.
    - Software: Training of SMC and teachers; creation and training of a school hygiene club; hygiene education and behaviour change activities in the schools and surrounding communities; establishing management systems to ensure sustained financing, monitoring and maintenance.
- Located in a secure region (schools located in Gao and Timbuktu regions as well as northern areas of Mopti region were excluded)

New beneficiary schools were selected to replace those from the baseline that were ineligible. Stratified selection was employed to ensure that the evaluation sample was representative of the overall program according to implementing partner, school size, and whether implementation was finalized before or after the start of the study period in January 2013. Schools were randomly selected from within each stratum, with probability proportional to size sampling, using a random number generator in Excel. All CARE-supported schools were in the Gao, Timbuktu, and Mopti regions, which were affected by the ongoing insecurity. Ten CARE-supported schools were included from the most secure part of Mopti region to retain partner participation without compromising study power in the event that those schools were lost to follow-up. This brought the total number of beneficiary schools in the evaluation up to 100. The distribution of primary schools within the DCIM program by partner and region is shown in Table 1, and number of beneficiary schools that participated in the evaluation are listed by partner and region in Table 2.

Table 1: Distribution of beneficiary primary schools

	CARE	OXFAM	SAVE	UNICEF	WATERAID	Total
Bamako	-	-	-	-	30	30 (4%)
Gao	-	25	-	-	-	25 (3%)
Koulikoro	-	75	-	105	15	195 (27%)
Mopti	100	-	-	89	-	189 (26%)
Sikasso	-	-	141	84	-	225 (31%)
Timbuktu	70	-	-	-	-	70 (9.5%)
Total	170 (23%)	100 (14%)	141 (19%)	278 (38%)	45 (6%)	734

Table 2: Distribution of beneficiary evaluation sample schools by region and partner

	CARE	OXFAM	SAVE	UNICEF	WATERAID	Total
Bamako	-	-	-	-	5	5
Koulikoro	-	15	-	27	3	45
Mopti	10	-	-	-	-	10
Sikasso	-	-	21	19	-	40
Total	10	15	21	46	8	100

Schools that had not participated in any substantial WASH programs in the past three years were identified from government records with the assistance of DCIM partner staff and education officials. These schools were matched with beneficiary schools enrolled in the evaluation in the same educational district (CAP) to serve as comparison schools. Matching criteria included school size and the presence of latrines and water points so that the WASH characteristics in the comparison sample matched the characteristics of the beneficiary sample at baseline. Nine of the comparison schools that had participated in the 2011 baseline survey were retained for the impact evaluation. The remaining baseline comparison schools were ineligible due to the reselection of the beneficiary school sample as well as a more stringent selection protocol.

Following two rounds of data collection, we conducted a simulation-based power calculation to estimate our power and estimate the number of follow-on periods required based on our revised sample size and observed absence and diarrhea rates. We found that we had 80% power to detect a 20% reduction in risk of absence and a 30% reduction in risk of diarrhea given 5 follow-up visits. Some schools were not eligible for analysis during the initial data collection rounds as they had not yet received the complete DCIM intervention; we added a sixth follow-up visit to minimize loss of power.

DATA COLLECTION

Rates of diarrhea and absenteeism are known to vary widely throughout the year. Data were therefore collected repeatedly at each school over the length of the evaluation period (longitudinal data collection) to allow the capture of inter-seasonal and inter-annual variations in impact indicators. Data were collected by trained enumerators, fluent in French, Bambara, Senoufou, and Pulaar, every 6-8 weeks between January 2013 and May 2014. Each school was visited 5-6 times for a total of 1096 school visits. A total of 887 visits post-implementation visits were included in the analysis. Data collection activities during these visits included:

- Interviews with school directors concerning the school WASH environment and school demographic information (1096 total interviews);
- Roll call of attendance in all classes using logs maintained by the enumerators with a roster of all pupils in grades 1-6, typically age 6-14, by class and by sex (1061 total roll calls);

- Observations of latrines, water points, hygiene materials, cleaning supplies, and other WASH hardware and facilities (1096 total observations);
- Observations of pupil handwashing practices after leaving latrines (153 total observations, with 2,604 total pupils observed).

Within each school, a sample of 20 boys and 20 girls from grades 3-6 were randomly selected from class registers. We interviewed students in grades 3-6 based on the ability of children at this grade level to reliably answer survey questions. This cohort of pupils was followed throughout the evaluation period. If a pupil in the cohort left the school during the evaluation period due to abandonment or transfer, that pupil was replaced by another randomly selected pupil of the same sex in the same class. Pupils in the 6th grade who advanced to secondary school at the end of the 2012-2013 school year were replaced by pupils in the 3rd grade at the start of the 2013-2014 school year. These students were interviewed at each school visit for a total of 10,068 pupils interviewed (47% girls) and 43,265 total interviews. Questions included:

- Recall of absence and reason for absence in the past 7 days;
- Recall of diarrhea and respiratory infection symptoms in the past 7 days;
- WASH practices at home and at school.

A 7-day recall period was used to ensure a sufficient reporting period while minimizing bias [2]. In some instances, a data collection activity may not have been conducted in all rounds, or data may not have been available in all schools. For each indicator in this report, results are presented for all observations for which we have data to evaluate the indicator. Survey instruments are included in Appendix A.

Finger-prick blood spots were also collected from 808 pupils in 20 beneficiary and 20 comparison schools between February-April 2014. Analysis of these samples is ongoing and results will be reported separately.

Enumerators participated in a one-week training that included data collection procedures, ethics, minimization of bias, and use of android devices and the ODK data collection software. Enumerators participated in pilot field trials and were asked to independently record answers on their devices. Answers were then compared between enumerators to identify areas where responses differed for further clarification and training. Refresher trainings on ethics, bias, and data collection methods along with field tests were conducted at the start of the second school year in October 2013.

Data were entered electronically on Android enabled phones using Open Data Kit (ODK) Collect software https://opendatakit.org/. Collected data were transmitted daily to an online server configured with ODK Aggregate software using the phones' mobile data connection. In the infrequent event that mobile data coverage was not available, data was stored on the phone and sent when the enumerator returned to an area with coverage. Data on the server was reviewed on a weekly basis to ensure that data was being entered correctly. New versions of the data collection forms were periodically uploaded to the server for the enumerators to download and employ.

OUTPUT AND OUTCOME ANALYSIS

Analysis of output and outcome indicators can be found in Appendix B. Data were analyzed using Stata Statistical Software: Release 13 (College Station, TX). The analysis for output and outcome indicators presented in this report is the average percentage of schools meeting each indicator. This analysis was conducted pre- and post-implementation, by intervention status and by partner; according to the number of months that had passed since implementation of the intervention was completed (time since implementation), by intervention status and by partner; and by location in rural and urban areas, among beneficiary schools only. Baseline data has been included for all DCIM beneficiary schools in the regions that participated in the impact evaluation (schools in Timbuktu and Gao excluded). Since few of the non-beneficiary schools surveyed at baseline were selected as matched controls for the impact evaluation, baseline data are not presented for comparison schools.

Independent samples t-tests were used to compare demographic information between the beneficiary evaluation sample and the beneficiary baseline and comparison evaluation sample. Chi-square tests and logistic and linear regression, adjusted for repeated observation at the school level, were employed to assess bivariate associations between observed WASH characteristics and school size and location in urban or rural areas. Independent sample t-tests were used to assess the relationship between governance indicators and observed WASH characteristics. These bivariate analyses were conducted only among beneficiary schools. They are exploratory analyses indented to identify possible associations between key WASH indicators and factors that program partners identified as potentially influencing WASH outcomes. Identifications of these associations does not imply causation, but can be used to help program partners understand and identify conditions that may require additional programming considerations in the future.

Analysis of impact indicators is described in the impact section below. Key indicators assessed in this evaluation are listed in Table 3.

ETHICS

This evaluation was approved by Emory University's Institutional Review Board (IRB). The Malian Ministry of Education requested that we seek authorization for school-based studies from the *Centre National de la Recherche Scientifique et Technique* (CNRST), or National Technical and Scientific Research Center. The Emory IRB, the CNRST, and the Ministry of Education approved consent in loco parentis (in the place of parents) signed by the school director and the school management committee (*Comité de Gestion Scolaire*), the organization empowered to oversee management and activities at the school, on behalf of the community that school serves. Pupils who were selected for the evaluation provided informed verbal assent. The evaluation was registered at ClinicalTrials.gov, identifier NCT01787058.

Table 3: Key impact, outcome, and output indicators

Type	Impact on…	Indicator	Data collection method
Impacts	Health	• Diarrhea	• 7-day pupil recall
		• Respiratory infections	• 7-day pupil recall
		• Enteric illnesses	• Analysis of antigens collected from blood spots
	Education	• Absenteeism	• 7-day pupil recall • All-school rollcall
		• Enrollment (general and girls) • Abandonment • Passing rate	• Review of school records
Outcomes	Improvement of hygiene practices	• Open defecation	• Observation of school grounds and pupil practice • Director report
		• Handwashing upon leaving latrines	• Observation at break time
	Reinforcement of governance / maintenance	• Functionality of CGS • Designation of responsibility for WASH • Spending on / budget for WASH • Regularity of latrine cleaning	• Director report
Outputs	Access to safe water	• Availability of a functional, improved water point • Availability of drinking water containers	• Director report • Observation
	Access to improved, safe, clean, gender-separated latrines	• Availability of latrines in good condition	• Observation
	Access to hygiene materials	• Availability of soap, detergent, drinking water and handwashing containers, anal cleansing kettles	• Observation

All pupil health information entered into the data collection devices was de-identified and the data collection devices were password protected. Paper logs linking pupils to their identification numbers were kept securely by the enumerators. No personal health information was collected from teachers. No compensation was offered or provided for participating in this evaluation. Comparison schools were informed that partner organizations would place them in high priority for receipt of future WASH programs. All interviews were conducted in a private but visible space to minimize the risk of sharing personal health information while maintaining safety. All efforts were made to minimize the amount of lost class time for participation in research activities.
While the finger-prick to collect dried blood spots may have caused momentary physical discomfort, there was no excess risk associated with this procedure. Funding for this study was provided by Dubai Cares through the five DCIM partners. The authors conducted this evaluation independently and do not declare any conflicts of interest.

B. SCHOOL DEMOGRAPHICS

A total of 100 DCIM beneficiary schools and 100 comparison schools participated in this evaluation. Data collection visits for the impact evaluation began prior to the completion of program activities in some beneficiary schools. Information collected during visits that occurred before all intervention components (water points, latrines, hygiene kits, and initial IEC training) were complete at that school is excluded from analysis. Due to time constraints, we could not visit all schools in the 6th data collection round; beneficiary schools that had not yet received the complete intervention during our first two rounds of data collection and their matched comparisons were prioritized. A breakdown of the number of beneficiary schools included in the analysis at each data collection round is shown in Table 4. As noted above, we included an additional data collection round to minimize the effect that this ineligibility for inclusion in analysis could have on our ability to detect impacts.

Table 4: Number of beneficiary schools included in analysis by data collection round

	Baseline Mar-Apr 2011	**Round 1** Jan-Mar 2013	**Round 2** Apr-May 2013	**Round 3** Oct-Dec 2013	**Round 4** Dec 2013-Feb 2014	**Round 5** Mar-Apr 2014	**Round 6** May 2014
CARE	80	4	4	4	10	10	10
OXFAM	75	6	12	14	15	15	10
SAVE	130	11	13	16	21	21	10
UNICEF	226	19	19	46	46	46	29
WATERAID	45	7	7	7	8	8	2
Total	**556**	**47**	**55**	**87**	**100**	**100**	**61**

Demographic information for DCIM beneficiary schools at baseline, DCIM beneficiary schools included in the evaluation, and comparison schools included in the evaluation are presented in Table 5. Enrollment data was taken from school records at baseline (Mar.-Apr. 2011) and from the fourth round of evaluation data collection (Dec. 2013-Feb 2014). Baseline latrine and water point information was taken from the baseline data for beneficiary schools and the first round of evaluation data collection for the comparison schools.

Enrollment and WASH characteristics were comparable between schools in the beneficiary evaluation sample and the overall DCIM program. The percentage of schools classified as urban was significantly higher in the evaluation sample.[1] Total average enrollment was slightly higher in the beneficiary evaluation sample compared to comparison schools, but the difference was not significant. Boys' enrollment was significantly higher, although girls' enrollment was not. Efforts were made to match comparison schools to beneficiary schools based on school size; however, the partners' strategy of prioritizing large schools for involvement in the program meant that

[1] Baseline classification of urban or rural zone was based on official districting, while the classification used in the evaluation was based on presence of electricity, banks, and administrative buildings.

Table 5: Demographic information

Indicator	Beneficiary evaluation sample		Beneficiary baseline			Comparison evaluation sample		
	N	mean(sd)/n(%)	N	mean(sd)/n(%)	p	N	mean (sd)/n(%)	p
Enrollment, total	*100*	342.9 (217.3)	*556*	329.0 (233.1)	0.58	*100*	286.3 (200.7)	0.06
Enrollment, boys	*100*	181.4 (115.3)	*556*	173.2 (127.5)	0.55	*100*	148.4 (128.4)	**0.03**
Enrollment, girls	*100*	161.5 (105.3)	*556*	155.8 (111.1)	0.63	*100*	137.8 (117.4)	0.11
Baseline latrines	*100*	86 (86.0%)	*530*	450 (84.9%)	0.78	*100*	86 (86.0%)	1.00
Baseline water points	*100*	51 (51.0%)	*532*	241 (45.3%)	0.29	*100*	45 (45.0%)	0.72
% urban	*100*	31 (31.0%)	*556*	95 (17.0%)	**<0.01**	*100*	33 (33.0%)	0.76

comparably-sized comparison schools were not available in all areas. Beneficiary and comparison schools were comparable in terms of presence of latrines and water points and location in rural or urban settings.

The ongoing conflict in the north of country displaced a number of people, and internally displaced persons (IDPs) moved in and out of the study area over the course of the study. We did not track whether schools had internally displaced persons; however, we did track whether schools had experienced exceptional changes in enrollment. Sixty-six percent of beneficiary schools and 54% of comparison schools reported exceptional increases or decreases in enrollment during the evaluation period.

C. OUTPUT AND OUTCOME RESULTS

The following narrative highlights key output and outcome findings from the evaluation. According to the partners' internal monitoring system, all hardware and software implementation activities had been completed in each school prior to time that the school began participating in the evaluation.[2] The results presented here therefore do not reflect what the DCIM partners provided to the schools, but rather the degree to which hardware was still in place and functional and behavior change and governance had been adopted by the schools at different time points after implementation.

Where appropriate, outcome and output data are presented according to the number of months that had passed since each school received the complete intervention package (time since implementation). This was done in order to illustrate the success of schools in meeting the indicators over time. However, each time point does not correlate to a uniform cohort of schools throughout the evaluation period, since schools received the intervention at different times. As such, any assessment of trend should be made with caution since different schools are captured at each time point. For instance, schools that received the complete intervention in January 2012, one year prior to the start of data collection, are not included in the outcome results for the period of 1-12 months after implementation; likewise, schools that did not receive the complete intervention until January 2014, one year after the start of data collection, are only included in the outcome results for the period of 1-6 months after implementation. The implementation

[2] In the areas where CARE operated, environmental conditions and budget constraints did not allow for water points to be constructed on all school grounds.

approach also evolved over the program period and differences between time points may reflect the effect of different implementation approaches. In addition, these data are organized by time since implementation rather than the time of year that the data was collected. This may also impact the interpretation of these findings, as, for example, seasonality may influence the presence of water in a way independent of months since implementation. These are key limitation of this approach.

The sample size of the impact evaluation was powered to provide an accurate analysis of the overall impact of the DCIM program on the health and education of students. It was not powered to provide a statistically significant analysis of these impacts for individual partners, nor was it designed to provide a representative overview of outcome and output results for individual partners. As such, all data stratified by partner is not statistically powered to detect differences between partners and is reported for reference only.

1. DCIM WASH STANDARDS

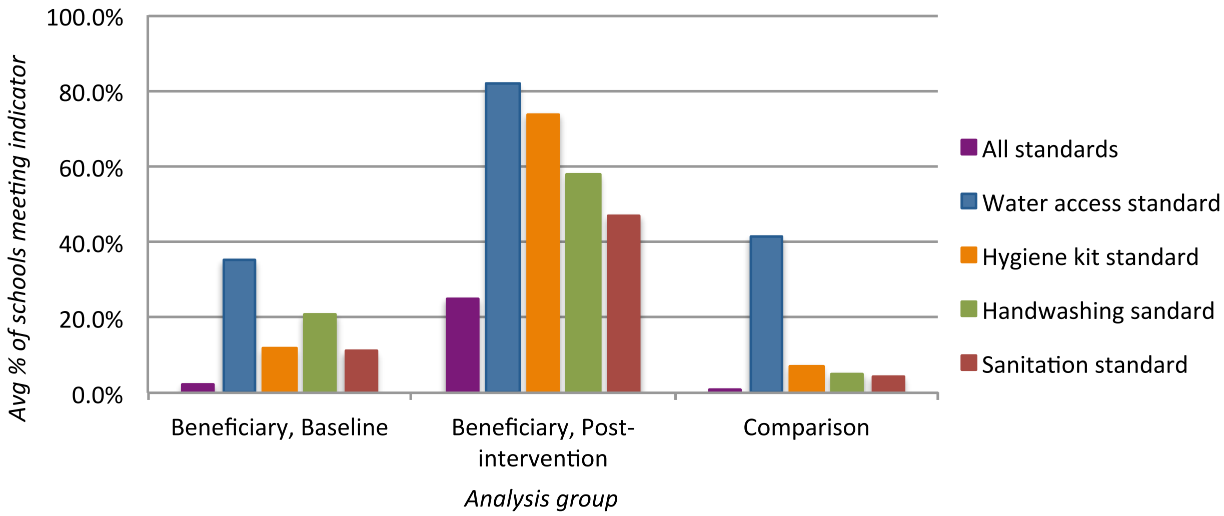
The DCIM partners established four school WASH standards related to water access, sanitation, handwashing, and hygiene kits. Standards were developed collaboratively by the implementing partners and Dubai Cares, and served as outcome-based implementation targets considered the minimum school conditions necessary to provide a safe and clean WASH environment for school children. The criteria included for each of these four standards are listed in Table 6 (15 criteria in total).

Table 6: DCIM WASH standards and criteria

Standard	Criteria
Water Access	• Water point is improved (tap, borehole well, or covered well) • Water point is located on the school grounds • Water point is functional (as defined by the school director)
Sanitation	• Latrines are located on the school grounds • Latrines are improved (cement slab) • Latrines are observed to be sex-separated in practice • For every 70 pupils, there should be at least one latrine that is: o Safe (no cracks or vibrations in floors, walls, or roof) o Clean (no visible signs of feces or pools of urine, the pit is not filled within 50cm of the top, observed to be "clean" or "somewhat clean")
Handwashing	• At least one handwashing container (including water kettles) available to pupils • Presence of water in at least one handwashing container • Presence of soap near at least one handwashing container
Hygiene Kits	• Presence of at least one drinking water container with a tap • Presence of at least one kettle for anal cleansing • Presence of handwashing soap • Presence of either bleach or detergent

During the first two rounds of impact evaluation data collection we did not collect sufficient information to calculate pupil:latrine ratios. Therefore, analysis of DCIM schools' compliance with the DCIM sanitation standard, as well as compliance with all four standards, is presented for data collected from October 2013 onward.

Figure 1: Achievement of all four DCIM school WASH standards



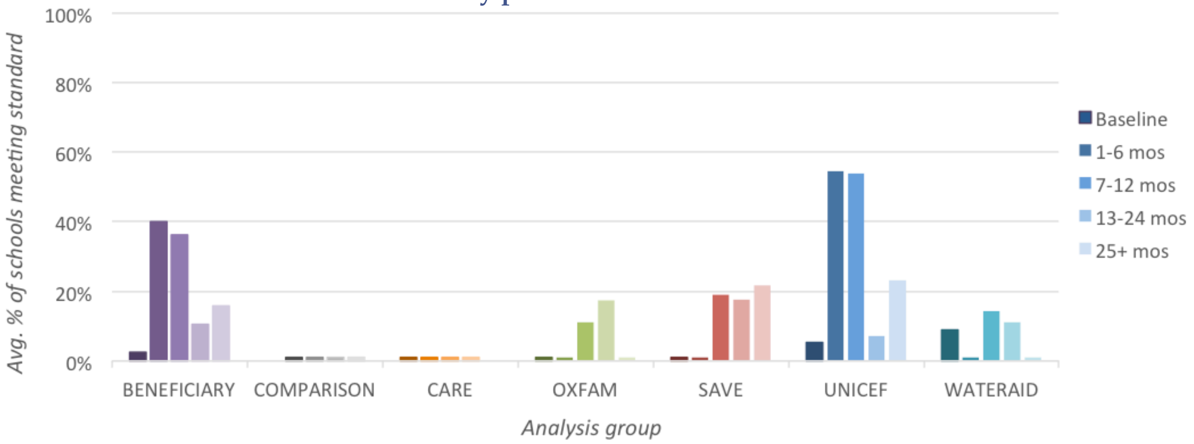
Overall, beneficiary schools were found to meet all four DCIM WASH standards during 24.9% of evaluation visits, compared to 2.5% of DCIM schools at baseline and none of the comparison schools (Figure 1). Achievement of the individual standards varied considerably. Beneficiary schools were most likely to meet the water access (81.8%) and hygiene kit (74.0%) standards, and least likely to meet the handwashing (57.9%) and sanitation (47.0%) standards. Beneficiary schools post-intervention performed better across all standards compared to both beneficiary schools at baseline and comparison schools.

Among beneficiary schools, there was no statistically significant difference between schools in urban or rural areas meeting all four DCIM standards (29.5% urban, 22.9% rural, p=0.19). Meeting all four indicators was not associated with school size (OR=1.00, p=0.59).

Figure 2 presents the achievement of all four DCIM standards overall and by partner, at baseline and by time since implementation.[3] The trend across all DCIM beneficiary schools (purple bars) indicates that DCIM beneficiary schools performed better than comparison schools at all time points. Beneficiary schools were most likely to meet all four DCIM standards during the first year following complete implementation of the intervention; schools visited 13 months or more after implementation showed much lower achievement of the standards.

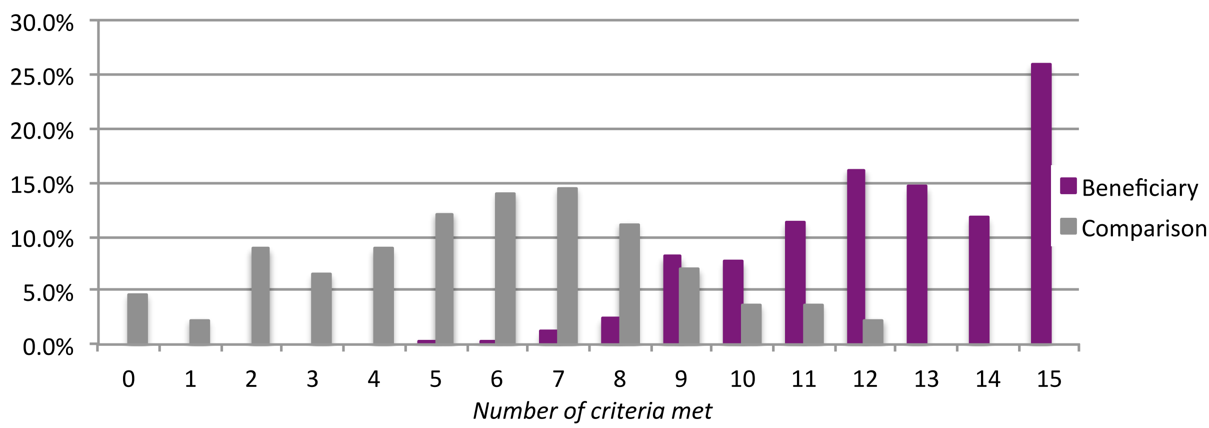
[3] There is no baseline data for comparison schools.

Figure 2: Achievement of all four DCIM school WASH standards by intervention status and by partner over time



The trend in the average percentage of schools meeting all four WASH standards was driven by UNICEF, as the intervention was managed by UNICEF in 46 out of the 100 beneficiary schools participating in this evaluation. The percentage of schools meeting all four standards for Oxfam, Save the Children, and WaterAid were lower overall compared to UNICEF, and none of the other partner schools met this standard in the time period 1-6 months after implementation.[4] Oxfam and WaterAid schools showed signs of lower rates of achievement as time since implementation increased, although Save the Children schools appeared to maintain rates over time. None of the ten CARE schools met all four standards at any time point, for the reasons mentioned above. The success of partners' achievements towards the standards could not be evaluated independently of the context, as each partner worked in settings that created unique challenges. However, each partner's approach was designed to accommodate the different contexts.

Figure 3: Percentage of observations where DCIM criteria were met



Schools were evaluated for fifteen criteria to determine whether they met the DCIM standards. Figure 3 presents the percent of observations when criteria were met for beneficiary and comparison schools. Although beneficiary schools met all fifteen criteria 25% of the time on average, they met thirteen or more criteria at 52% of

[4] Individual partner data may be artificially influence by small sample sizes

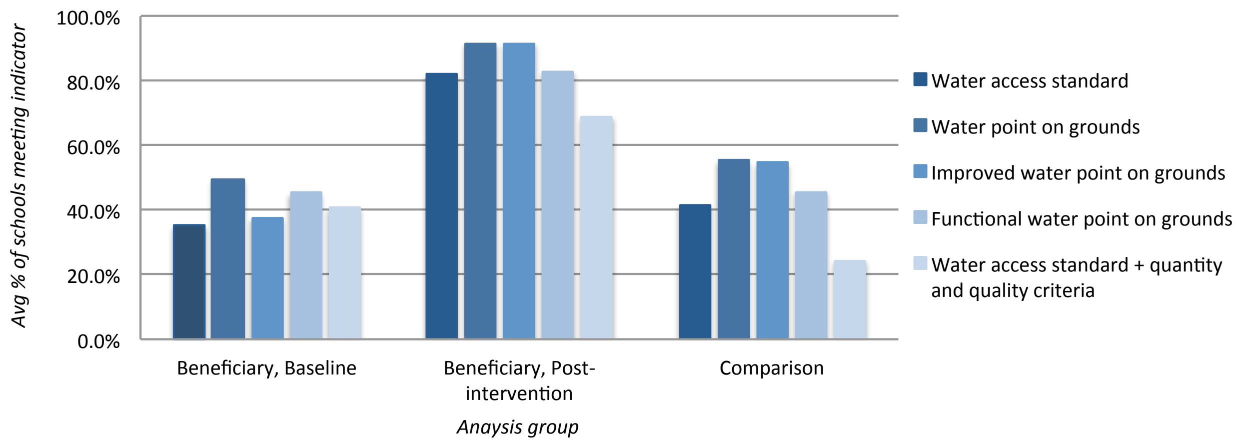
observations, and beneficiary schools met ten or more criteria at 87% of observations. Comparison schools met six or fewer criteria at 58% of observations and no comparison schools met more than twelve of the criteria at any time.

The following sections report in detail on each of the four DCIM WASH standards and their criteria.

WATER ACCESS

The DCIM standard for water access is the presence of an improved and functional water point in the school compound. Presence of a water point was based on interview with the schools director and confirmed by structured observation. Functionality was assessed by report of the schools director. Taps, borehole wells, and covered wells were classified as improved water points.

Figure 4: Percentage of schools meeting DCIM water access standard and criteria



Overall, DCIM program schools met the water access standard at 81.8% of observations, compared to 35.7% at baseline and nearly double that of comparison schools (41.1%) (Figure 4). An improved water point was present on the school grounds at nearly all beneficiary schools with the exception of some schools overseen by CARE.[5] Improved water points included taps or stand pipes (13 schools), borewells, (77 schools), and covered wells (18 schools). Beneficiary schools were observed to a functional improved water point on the schools grounds at 81.8% of observations.

Water points were also assessed for whether they met more stringent quantity and quality criteria, including being more than 15m from any latrine and providing a sufficient quantity of good-quality water throughout the school day and throughout the entire school year, according to the report of the school director. DCIM schools met these criteria 68.9% of the time on average, an improvement from 40.7% at baseline and 24.5% average among comparison schools (Figure 4).

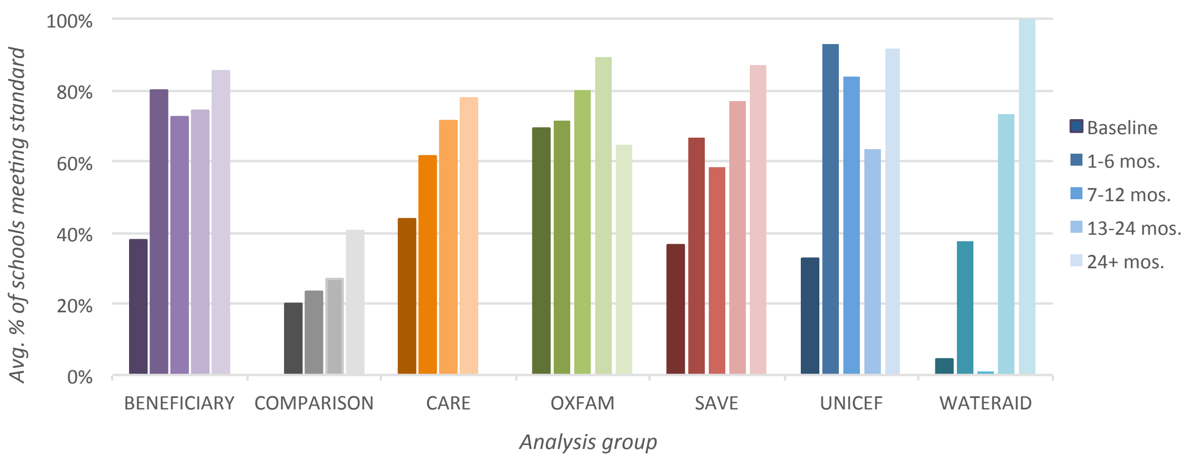
[5] As stated previously, in the areas where CARE operated, environmental conditions and budget constraints did not allow for water points to be constructed on all school grounds. Schools without water points were provided with water storage cisterns to support them in meeting their water needs.

Beneficiary schools in urban areas were more likely to meet the water access standard than were schools in rural areas (93.0% urban, 76.6% rural, p<0.01). Larger beneficiary schools were marginally more likely to meet the standards as compared to smaller schools (OR=1.004, p=0.07). Among schools with improved water points on the school grounds, there was no significant association between the percentage of water points that were functional in urban and rural areas (93.7% urban, 90.2% rural, p=0.24). Nor was there any significant association between the percentage of water points that were functional among schools that did or did not share their water points with the community (98.3% shared, 100.0% not shared, p=0.34). There was no association between school population and functionality of water points (OR=1.000, p=0.95).

Although improved and functional water points were available on the schools grounds at 81.1% of observations, water was available for pupils to drink at beneficiary schools during 94.9% of observations, compared to 62.6% of observations at comparison schools. This water was brought in from improved and unimproved sources off of the school grounds.

Figure 5 presents the percentage of schools meeting the DCIM water access standard overall and by partner, at baseline and by time since implementation.[6] There was some variation in the overall average percentage of schools with a functional water point, but we saw little evidence that achievement of the water standards declined as time since implementation increased.

Figure 5: Achievement of DCIM water access standard by intervention status and by partner over time



All WaterAid schools, which were located in urban areas and typically connected to a piped water supply network, had functional improved water points at all evaluation data collection rounds. Water points at schools overseen by partners excluding CARE were observed to be functional at 80-89% of visits and appeared to remain constant over time with minor exceptions.

---

[6] There were no CARE schools where program activities had been complete more than 24 months prior in the evaluation sample.

Figure 6: Percentage of beneficiary schools meeting water access criteria over time

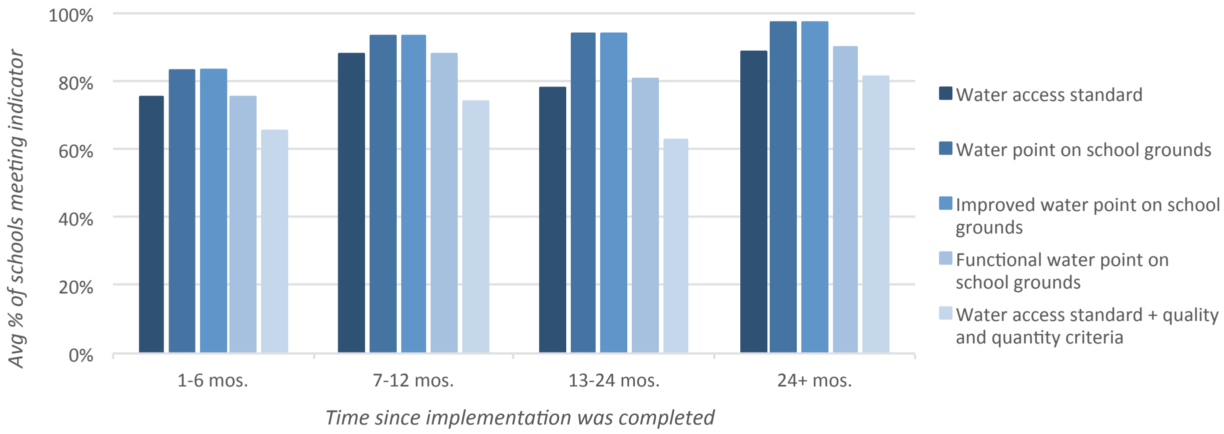


Figure 6 presents the percentage of beneficiary schools that met each of the criteria within the water access standard over time since the provision of the full intervention package. There did not appear to be any large changes in the percentage of schools that met each of the criteria as time since implementation increased.

SANITATION

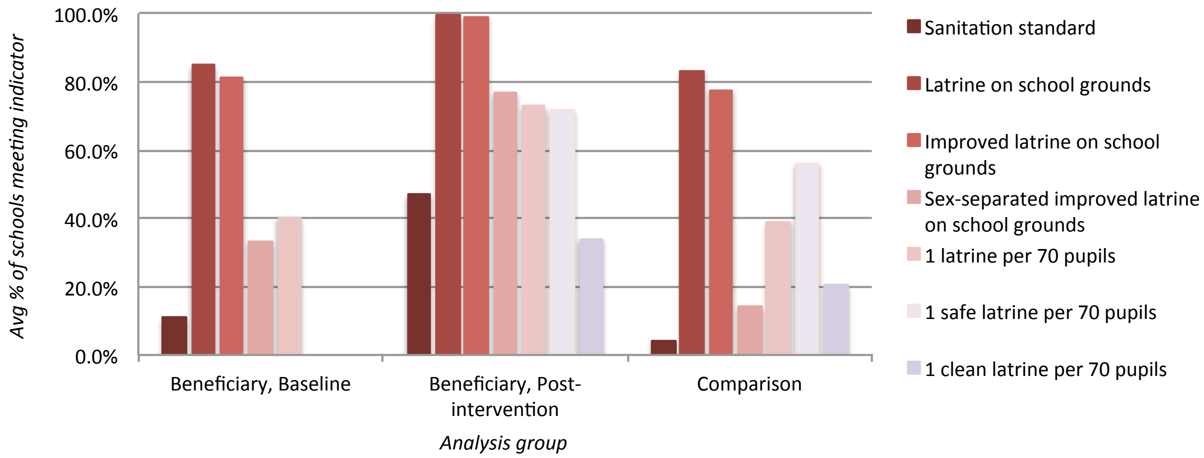
The DCIM standard for sanitation was the presence of at least one improved, sex-separated latrine on the school grounds that was clean and safe per 70 pupils. Latrines were individually assessed through structured observations. Latrines were assessed for being sex-separated through observation of use patterns during school breaks. They were considered improved if the floor was a cement slab. They were considered clean if there were no traces of feces or pools of urine, the pit was not filled within 50cm of the top, and they were classified as "clean" or "somewhat clean" by the enumerator. They were considered safe if there were no large cracks in floors, walls, or roof. The pupil:latrine ratio was calculated by dividing the number of latrines that met each criteria by the total school population. Latrines were not included in the calculation if they were locked and therefore not accessible to pupils.

Of the four DCIM WASH standards, schools were least likely to meet the sanitation standard; 47.0% of beneficiary schools meet this standard on average across all observations. This is an improvement compared to baseline (11.6%) and comparison schools (4.4%) (Figure 7).

Improved latrines were available to pupils at 99.1% of observations in beneficiary schools, compared to 79.9% at baseline and 77.6% of observations in comparison schools. Latrines were observed to be sex-separated in practice at 76.8% of observations. Program schools achieved a mean ratio of 60 pupils per latrine, compared to 114 pupils per latrine in comparison schools and 110 pupils per latrine at baseline. On average 72.9% of beneficiary schools provided one latrine per 70 pupils and 71.5% of schools provided one latrine that met the program's safety standard. Beneficiary schools provided one clean latrine for every 70 pupils at 56.1% of observations.[7]

[7] Baseline data not available for the number of latrines meeting safety and cleanliness criteria.

Figure 7: Percentage of schools meeting DCIM sanitation standard and criteria



There was no difference in the percentage of beneficiary schools in urban and rural areas that met the overall DCIM sanitation standard (47.6% urban, 46.7% rural, p=0.88), although schools were less likely to achieve the standard as enrollment size increased (OR=0.99, p<0.01). Urban schools were more likely to ensure that latrines were sex-separated (86.7% urban, 72.3% rural, p=0.04). Smaller schools were significantly more likely to meet overall ratio (1 latrine per 70 pupils: OR=0.98, p<0.01), cleanliness (OR=0.99, p<0.01), and safety criteria than larger schools (OR=0.99, p<0.01). There was no difference between urban and rural schools in meeting the overall ratio, safety, or cleanliness criteria, and there was no association between school size and sex-separation of the latrines.

Figure 8: Achievement of DCIM sanitation standard by intervention status and by partner over time

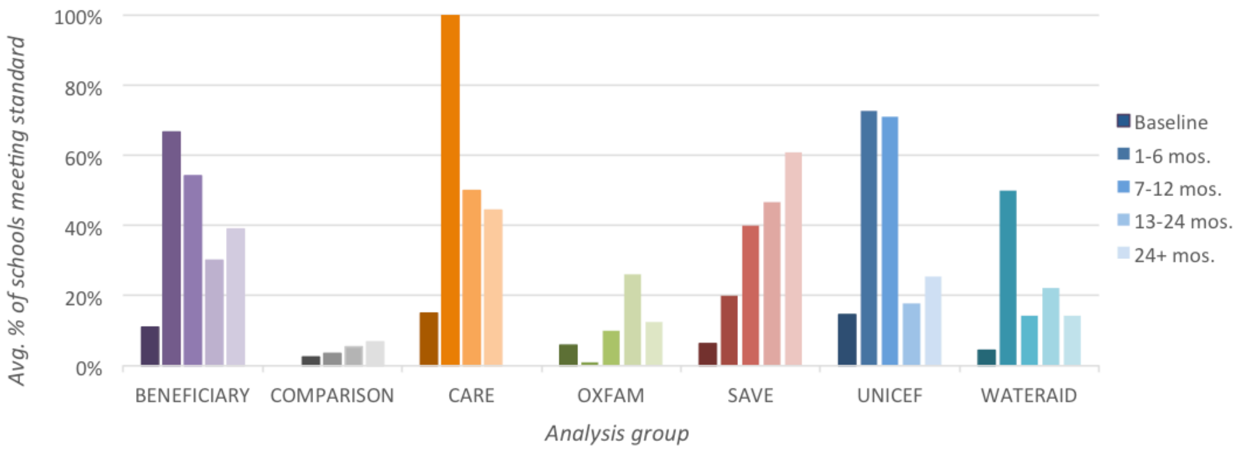


Figure 8 presents the average percentage of schools meeting the sanitation standard at baseline and by time since the schools received the full DCIM intervention package, by intervention status and by partner. Overall, DCIM beneficiary schools performed the best during the period of 1-6 months after implementation of the complete program package. As the time since provision of the intervention increased, schools tended not to succeed as well in meeting the sanitation standard.

Schools overseen by CARE, UNICEF, and WaterAid followed the overall trend; they were most likely to meet the sanitation standard during the first six months following provision of the intervention but the percentage of schools meeting the standard was lower after the 1-6 month period (in the case of UNICEF, after the 7-12 month period). Schools overseen by Save the Children and Oxfam fared better with increased time since provision of the intervention.

Figure 9: Percentage of beneficiary schools meeting sanitation criteria over time

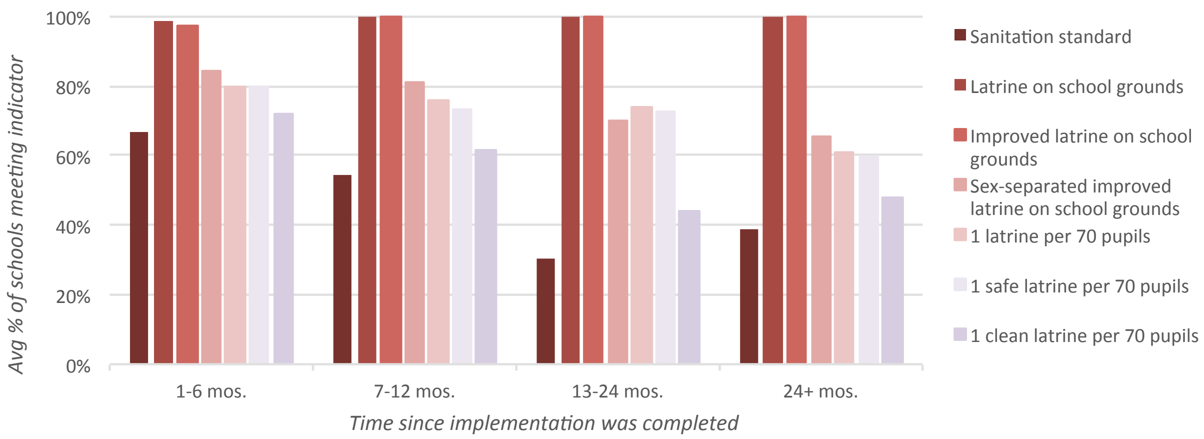


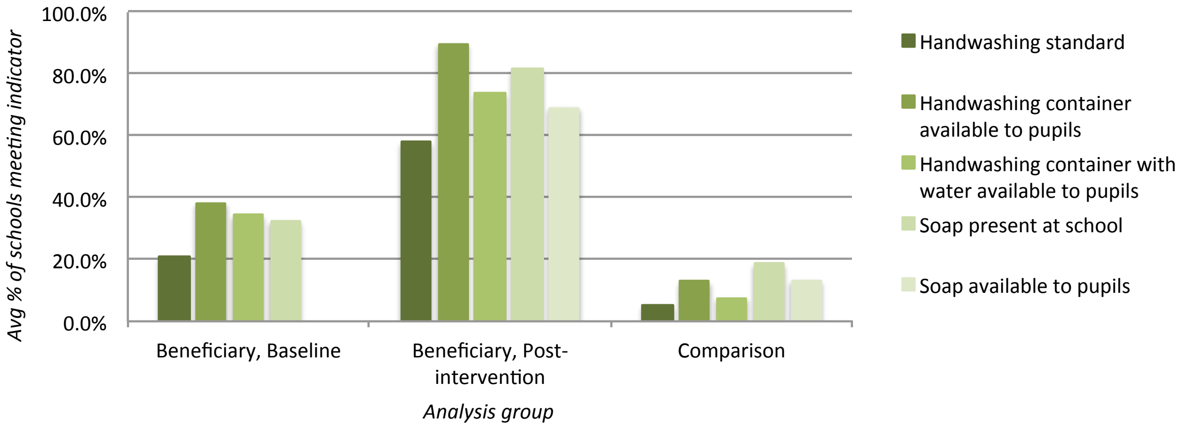
Figure 9 shows how all DCIM beneficiary schools fared at meeting each of the criteria within the sanitation standard over time since implementation. The percentage of schools with at least one improved latrine on the schools grounds remained constant over time. As time since provision of the intervention increased, schools were less likely to enforce separation of the latrines by sex or maintain a sufficient number of safe and clean latrines. Maintaining clean latrines was the criteria that beneficiary schools were least likely to meet at all time points.

HANDWASHING

The DCIM standard for handwashing was the presence of at least one handwashing container with water and soap. The presence of handwashing containers and soap was assessed through structured observation. Anal cleansing kettles were included as handwashing containers if enumerators observed pupils to use them for that purpose.

Overall, DCIM beneficiary schools met this standard at 57.9% of observations after provision of the intervention package, compared to 19.9% at baseline and 5.2% of comparison schools on average (Figure 10). At least one handwashing container was present in beneficiary schools at 89.8% of observations on average, and water was available in at least one of the containers at 73.7% of observations. The criteria that beneficiary schools were least likely to meet was the availability of soap near the handwashing stations. Although soap was observed to be present at the school 81.6% of the time on average, soap was observed to be available to pupils at 69.2% of observations.

Figure 10: Percentage of schools meeting the DCIM handwashing standard and criteria



Schools were also evaluated for meeting elevated handwashing standards, which included having at least one handwashing container (defined as a container with tap or tippy tap, but not kettles) with soap and water for each block of latrines. Beneficiary schools met this criteria at 42.9% of observations, compared to 1.2% of comparison school observations. As a proxy for handwashing practice, we also observed whether there was at least one handwashing station with signs of use. On average, 72.8% of beneficiary schools met this standard during the evaluation period, compared to 6.5% at baseline and 5.2% of comparison schools.

Cement handwashing stations were present in 54 schools. The proportion of cement containers observed to be functional and in use dropped sharply as time since implementation increased, from an average of 44.8% in the first six months after implementation to 16.0% one year after implementation and 0.0% two years after implementation. Similar data was not available for other types of handwashing containers, as broken containers may have been removed from the school.

Rural schools were more likely than urban schools to meet the handwashing standard (50.7% urban, 61.2% rural, $p=0.04$). There was no association between school size and meeting the standard ($OR=1.00$, $p=0.75$). Rural schools were more likely than urban schools to have water available in handwashing containers (urban 63.1%, rural 78.6%, $p<0.01$). There was no difference between urban and rural schools in the availability of handwashing containers (urban 86.2%, rural 91.2%, $p=0.13$) or whether soap was available to pupils (urban 70.6%, rural 66.2%, $p=0.35$), and no association between any of the handwashing criteria and school size.

Figure 11 presents the percentage of schools that met the handwashing standard by time since implementation. Overall, beneficiary schools were slightly less likely to achieve this standard as time since implementation increased. At 1-6 months after implementation 67.4% of schools met the standard on average, while 57.3% of schools met the standard at 2+ years after implementation. Schools overseen by CARE, Oxfam, and Save the Children appeared to perform better as time since implementation increased, while the percentage of schools managed by UNICEF and WaterAid meeting the criteria was lower with increased time since implementation.

Figure 11: Achievement of DCIM handwashing standard by intervention status and by partner over time

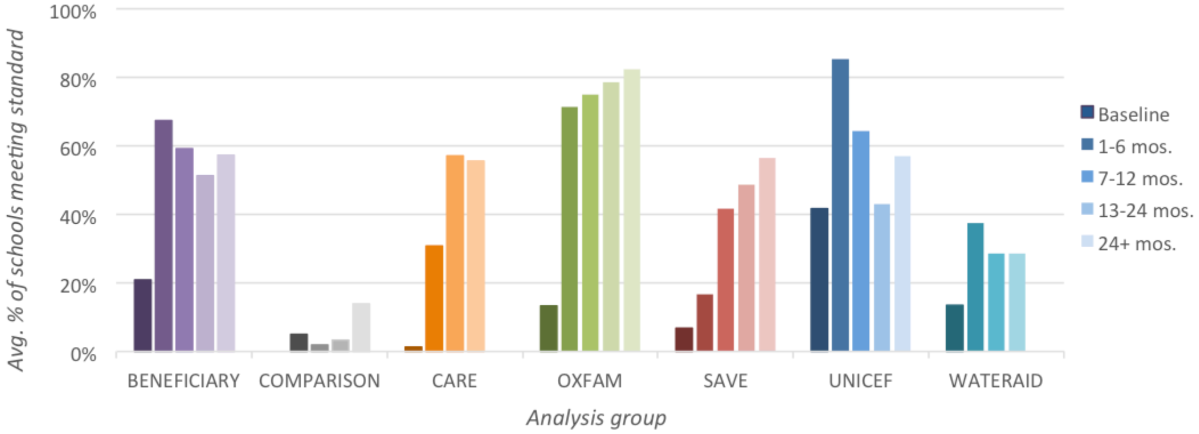
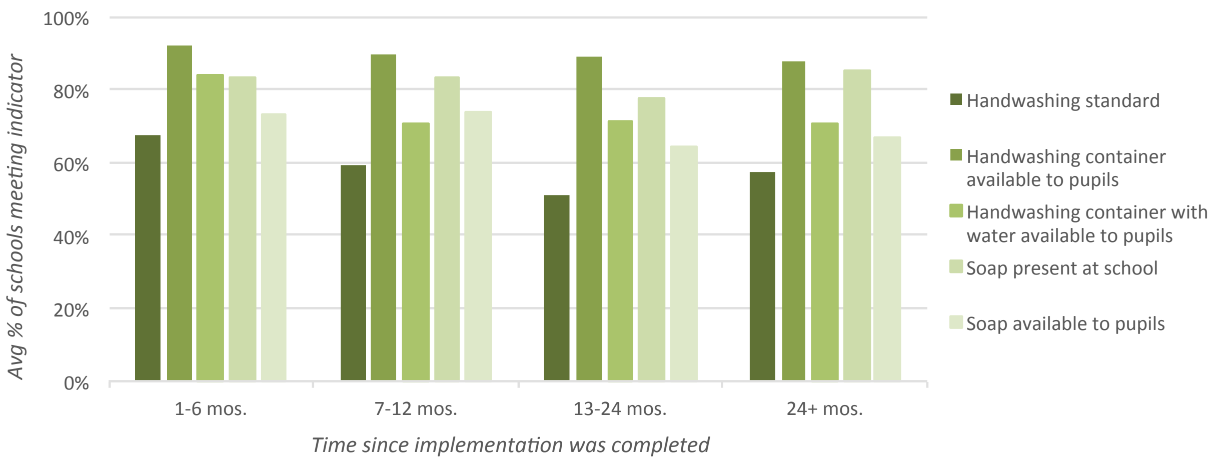


Figure 12 shows a breakdown of how all program schools fared at achieving each of the elements of the hygiene standard over time. Schools were slightly less likely to have handwashing containers present at school as time since implementation increased, although there was a larger decline in the availability of water within the containers over time. The percentage of schools with soap present at school remained constant as time since implementation increased, although on average fewer schools ensured that soap was available to pupils after six months following provision of the intervention package.

Figure 12: Percentage of beneficiary schools meeting handwashing criteria over time



HYGIENE KITS

The DCIM standard for hygiene kits is the presence of soap, detergent or bleach, at least one anal cleansing kettle, and at least one functional drinking water container. This was assessed through structured observation.

Overall, DCIM schools met this standard at 70.4% of observations, compared to 11.7% at baseline and 7.3% of observations in comparison schools (Figure 14). Beneficiary schools had drinking water containers and anal cleansing kettles at nearly all observations (water containers 97.9%, kettles 96.4%). Schools were less likely to have detergent or bleach (84.7%) or soap (81.6%). Availability of WASH supplies was not associated with school size (OR=1.000, p=0.68) or location in urban/rural areas (urban 68.3%, rural 74.9%, p=0.17).

Figure 14: Percentage of schools meeting DCIM hygiene kit standard and criteria

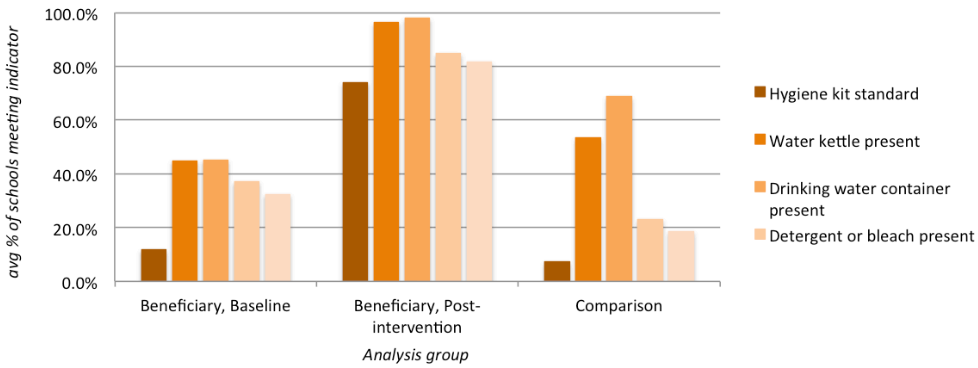


Figure 15 presents the percentage of schools that met the hygiene kit standard by the time elapsed since the schools received the complete intervention package. Schools were slightly less likely to achieve this standard as time since implementation increased. Overall, 82.5% of schools at 1-6 months after implementation met the standard, while 71.8% of schools met the standard at 2+ years after implementation. UNICEF and Oxfam schools performed the best at achieving this standard during the period of 1-6 months after implementation, although their schools did not perform as well as time since implementation increased. The average percentage of Save the Children and CARE schools that met the hygiene kit criteria was lower in the period of 1-6 months after implementation, but these schools performed better as time since implementation increased. WaterAid schools performed best at meeting the hygiene kit indicator during the period of 6-12 months following implementation, but schools that had received the intervention one or more years prior showed lower performance.

Figure 15: Achievement of DCIM hygiene kit standard by intervention status and by partner over time

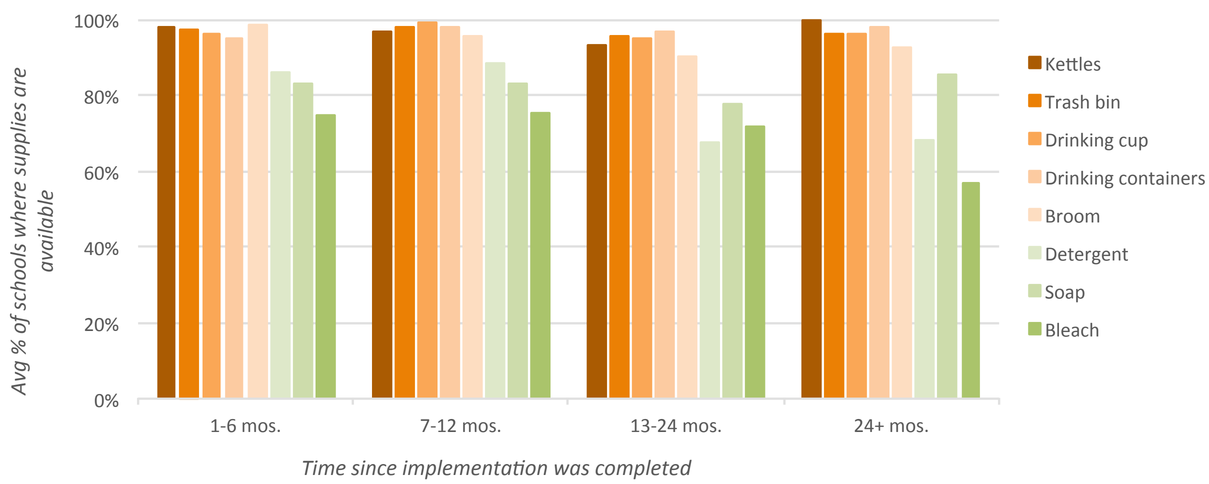
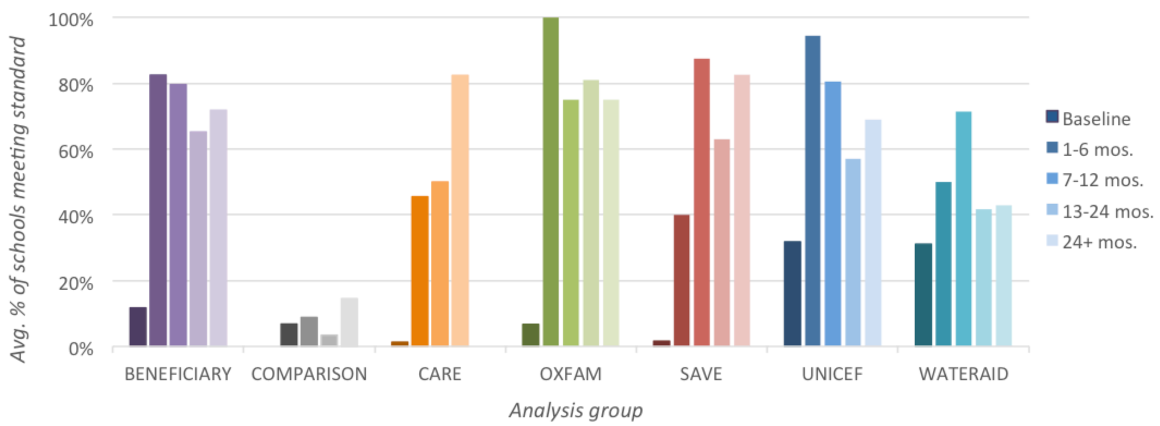


Figure 16 shows the presence of WASH supplies over time for all beneficiary schools. There were no large declines in the presence of durable materials such as kettles, trash bins, or water containers. However, schools were less likely to have detergent and bleach as the time since implementation increased. The presence of soap over time was more consistent.

Figure 16: Percentage of beneficiary schools with WASH supplies by time since implementation



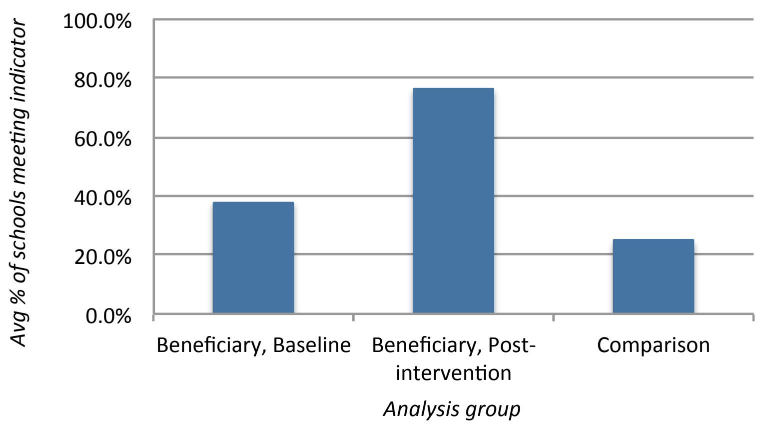
2. BEHAVIOR CHANGE

We tracked two outcome indicators related to pupil behavior change: open defection and handwashing.

OPEN DEFECATION

The cessation of open defecation was measured by both observation of the school grounds and report of the school director. The percentage of beneficiary schools without either evidence or report of open defecation improved from 38.1% at baseline to 76.2% of observations after provision of the intervention package (Figure 16). Open defecation was less likely to be practiced in beneficiary schools compared to comparison schools, where no report or observation of open defecation was found at 24.7% of observations.

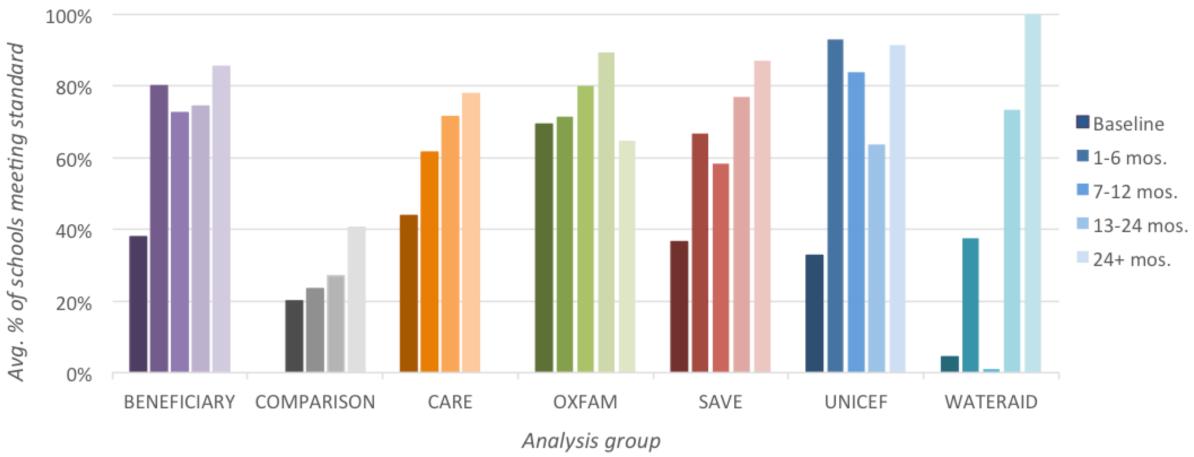
Figure 16: Percentage of schools without evidence or report of open defecation



There was no difference in the percentage of beneficiary schools without evidence or report of open defecation between urban and rural areas (urban 77.5%, rural 75.6%, p=0.67), and there was no association with school size (OR=1.001, p=0.60). Beneficiary schools with lower pupil:latrine ratios (OR=0.986, p=0.02) and those with a higher percentage of clean latrines (OR=9.376, p<0.01) were less likely to have evidence or report of open defecation. Among beneficiary schools, evidence of open defecation was less likely to be found when latrines were sex-separated (separated 84.9%, not separated 55.7%, p<0.01), when schools had at least one latrine per 70 pupils (>1 latrine per 70 pupils 80.8%, <1 latrines per 70 pupils 69.6%, p=0.02), when latrines met cleanliness criteria (>1 clean latrine per 70 pupils 65.6%, <1 clean latrine per 70 pupils 87.4%, p<0.01), and when schools met the overall DCIM sanitation standard (met standard 89.8%, did not meet standard 67.4%, p<0.01).

Results by time since implementation are shown in Figure 17. The overall percentage of schools where no open defecation was observed appeared to be constant over time, with the lowest average rates of open defecation seen among beneficiary schools in the period 24+ month after implementation. Among schools overseen by CARE, Save the Children, and WaterAid, the average percentage of schools without evidence of open defecation appeared to be greater as the length of time since implementation increased (the variability in WaterAid schools is due in part to the small sample size). UNICEF schools performed best in the 1-6 month and 25+ month periods. Oxfam schools started out with the lowest rates of open defecation and schools performed better as time since implementation increased until the period of 25+ months since implementation.

Figure 17: Percentage of school with no evidence or report of open defecation by intervention status and by partner over time



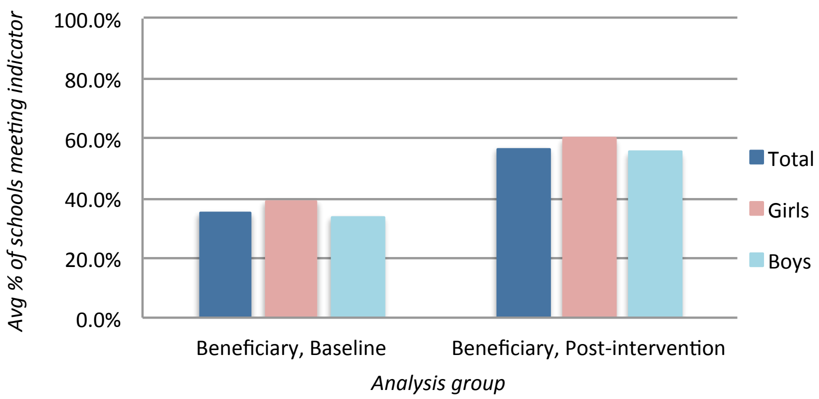
HANDWASHING

Pupil handwashing practices were observed at schools where water and soap were available. Enumerators conducted handwashing observations by stationing themselves in a place where they could observe both the latrines and the handwashing containers. Enumerators tracked pupils as they left the latrines during the morning or afternoon break and observed whether each pupil washing their hands with soap, washed their hands with water only, or did not wash their hands. Handwashing observations were not conducted if the school did not offer soap or if conditions were not suitable for being able to track pupils to the handwashing stations after they left the latrines.

At baseline, handwashing observations were conducted in one or fewer schools overseen by CARE, Oxfam, and Save the Children. Baseline data is therefore shown for the entire program only and not by partner. For comparison schools, the conditions for conducting handwashing observations were met on only five occasions during the evaluation period; due to this small sample size, results are not presented.

Handwashing observations improved from baseline. Overall, 57.7% of pupils in beneficiary schools where soap was available were observed to wash their hands with soap upon leaving the latrine, compared to 39.6% at baseline (Figure 18). Girls in beneficiary schools were more likely than boys to wash their hands with soap (girls 60.5%, boys 56.3%, p=0.05).

Figure 18: Percentage of pupils observed to wash their hands with soap upon leaving the latrine



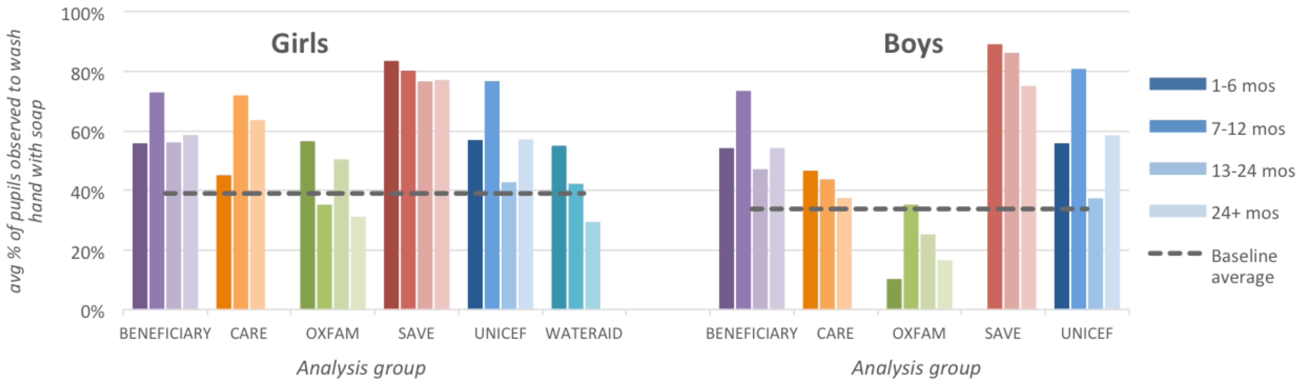
There was no difference in handwashing with soap among urban and rural beneficiary schools (urban 56.7%, rural 56.7%, p=1.00) and there was no association with school size (β=0.000, p=0.31). Pupils were more likely to wash their hands with soap if there was at least one handwashing container with soap per latrine block compared to those with less than one per block (66.3% 1≥container/block, 48.2% <1 container/block, p=0.05). Handwashing practice was not associated with total pupil:handwashing station ratio (β=0.000, p=0.16).

Figure 19 shows the percentage of girls and boys who were observed to wash their hands with soap by time since implementation.[8] As baseline data was not available for all partners, overall baseline average is presented for comparison. Handwashing practices among beneficiary schools were variable over time. Schools in the 7-12 month period saw the highest rates of handwashing with soap. Handwashing rates were alsovariable between partners. Although rates of handwashing for most partners

[8] No handwashing observations were conducted for boys in WaterAid schools during the evaluation period

improved overall compared to baseline, rates among Oxfam schools did not appear to improve. Rates among CARE schools were unusually elevated at baseline and dropped during the evaluation period.

Figure 19: Percentage of pupils at schools with soap available observed to wash hands with soap upon exiting the latrines, by intervention status and by partner over time



3. GOVERNANCE

This evaluation tracked several indicators for governance over WASH systems at schools, including:

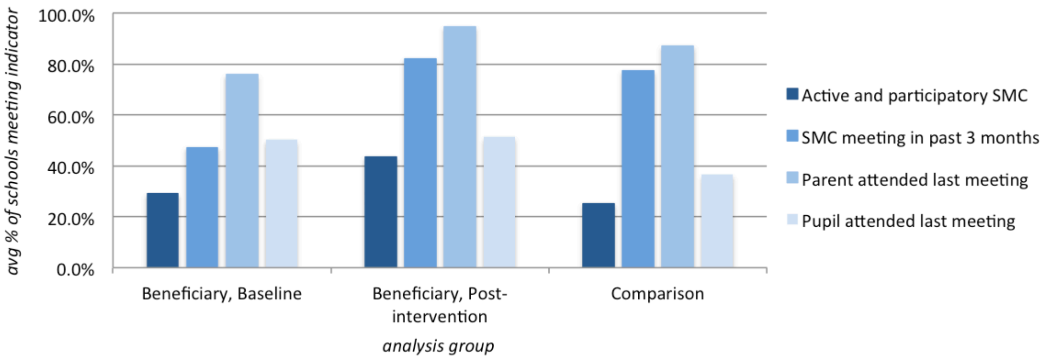
- An active and participatory school management committee (SMC), meaning at least one meeting in the past three months with the participation of parents and students;
- The existence of a budget dedicated to WASH;
- Funds being spent on WASH;
- Designation of teachers, students and parents to be responsible for overseeing WASH;
- Hygiene lessons taught in each class; and
- Frequency of latrine cleaning.

The rationale behind these indicators is that an appropriate system of operation and maintenance, financial management, and monitoring are assumed to contribute significantly to the success and sustainability of WASH in schools programs. Hygiene lessons in classes are assumed to provoke behavior change among pupils and empower them as agent of change in their community. We examined the relationship between governance indicators and key outcomes through bivariate analysis.

SCHOOL MANAGEMENT COMMITTEE

The criteria for an active and participatory school management committee was having at least one meeting in the past three months with the participation of parents and students. The percentage of beneficiary schools that met this criteria improved from 28.0% at baseline to 43.5% on average during the evaluation period, and beneficiary schools performed better compared to comparison schools (25.3%) (Figure 20). On average, parents were present at the last SMC meeting 94.6% of the time, and 82.2% of SMCs had met at least once in the past three months. Among stakeholders we tracked, pupils had the lowest attendance rate at SMC meetings (51.4%). The level of pupil participation was lower than at baseline (54.1%) though improved compared to comparison schools (36.5%).

Figure 20: Percentage of schools with an active and participatory SMC



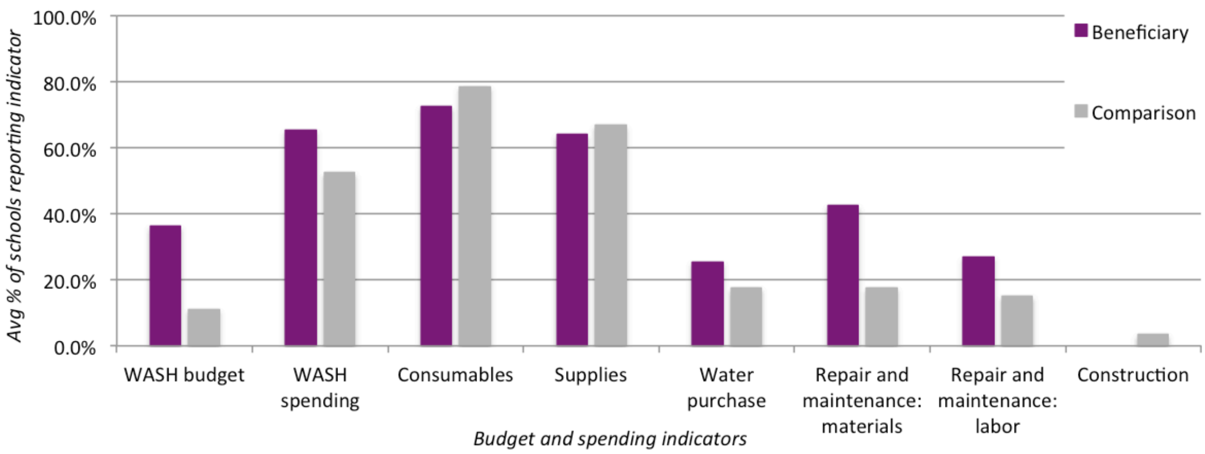
There was no relationship between an active and participatory SMC and school size (OR=1.00, p=0.74). There was also no difference in the average percentage of schools that met this criterion in urban or rural areas (urban 42.2%, rural 44.1%, p=0.80). There was no association between having an active and participatory SMC and meeting the latrine standard, the handwashing standard, and the hygiene kit standard. Surprisingly, schools that met the criteria for an active and participatory CGS were less likely to meet the DCIM water access standard than those that did not meet the criteria (active SMC 68.9%, non-active 88.9%, p<0.01).

WASH BUDGET

Figure 21 shows the percentage of schools that have a WASH budget, that spent money on WASH, and how the money was spent during the evaluation period. Comparable data was not collected at baseline. 36.3% of beneficiary schools reported a budget dedicated to WASH spending, although beneficiary schools reported having spent funds on WASH in the past three months at 65.4% of observations. Beneficiary schools were more likely to spend money on WASH overall than were comparison schools (52.4%) and were more likely to report having a WASH budget than comparison schools (8.6%). Beneficiary schools spent money on repairs and maintenance at higher rates than did comparison schools, while comparison schools were more likely to spend money on consumables and construction than beneficiary schools.

Beneficiary schools in urban zones were significantly more likely than schools in rural zones to report having spent money on WASH (urban 88.8%, rural 54.2%, p<0.01) and

Figure 21: WASH budget and spending by intervention status

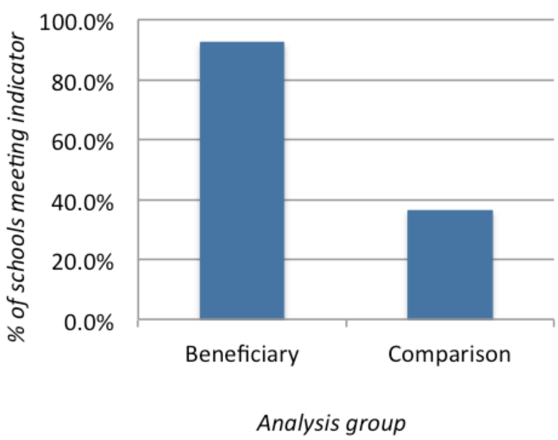


to have a budget for WASH (urban 42.7%, rural 19.2%, p<0.01). Larger schools were also more likely to report spending money on WASH (OR=1.007, p<0.01) and having a budget for WASH (OR=1.007, p<0.01). Schools that reported having a budget in place for WASH were more likely to meet the water access (budget 92.9%, no budget 77.3%, p<0.01), handwashing (budget 72.5%, no budget 56.0%, p=0.02), and hygiene kit (budget 84.1%, no budget 67.4%, p=0.01) standards. Schools that reported having spent money on WASH were more likely to meet the water access standard (spending 89.3%, no spending 69.2%, p<0.01).

DESIGNATION OF RESPONSIBILITY

Beneficiary schools reported that teachers, pupils, or SMC members were assigned responsibility of WASH at 92.3% of observations during the evaluation period, compared to 36.3% among comparison schools (Figure 22). Comparable data was not collected at baseline. Urban schools performed better than their rural counterparts (98.4% urban, 89.6% rural, p=0.03). There was no association between school size and designation of responsibility for WASH (OR=1.003, p=0.23). Schools with a teacher, parent, or student designated as responsible for WASH were more likely to meet the sanitation standard (designation 51.9%, no designation 13.3%, p<0.01). Designation of responsibility was not associated with any of the other three standards.

Figure 22: Percentage of schools where a teacher, pupil, or parent has been designated as responsible for WASH

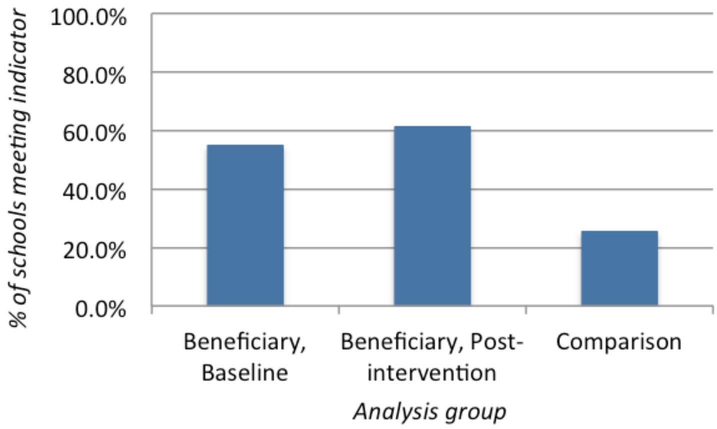


WASH LESSONS

The implementation of WASH curriculum was a core component of the program's information, education and communication (IEC) strategy. School director were asked at baseline (Mar-Apr 2011) and at the fourth evaluation data collection round (Dec-Feb 2014) whether a WASH lesson had been taught to all classes in the schools since the start of the school year. The content of these sessions was not evaluated.

WASH lessons were reported to have been taught in all classes at 61.0% of observations in beneficiary schools, compared to 54.7% of schools at baseline and 25.4% of observations in comparison schools (Figure 23). WASH classes were more likely to have been taught in urban beneficiary schools compared to rural schools, but the difference was not significant (70.0% urban, 57.4% rural, p=0.17). WASH classes were not associated with school size (OR=1.002, p=0.24). There was no difference in rates of handwashing with soap among schools where WASH classes had and had not been taught (WASH class 61.4%, no class 57.8%, p=0.77), nor were there significant differences in open defecation (WASH class 83.7%, no class 72.7%, p=0.12).

Figure 23: Percentage of schools where a WASH lesson has been taught in all classes since the start of the year

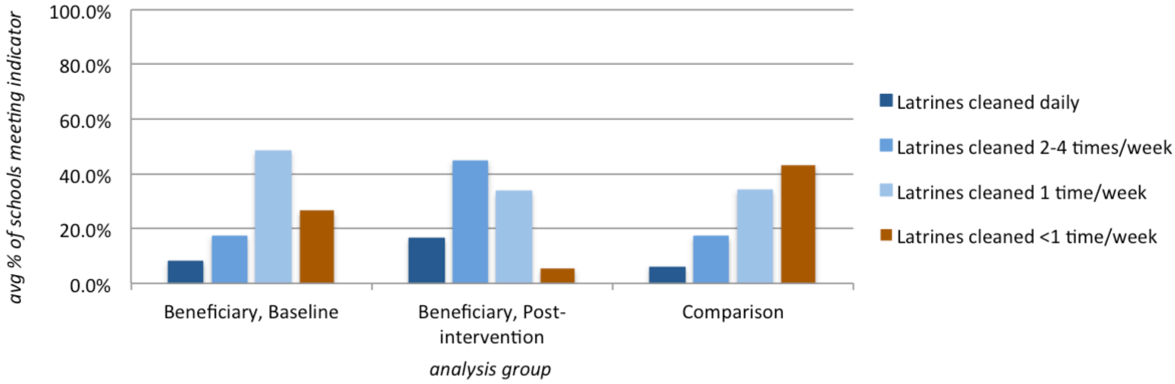


LATRINE CLEANING

Frequency of latrine cleaning was assessed through report of the school director. At baseline, 26.5% of beneficiary schools reported cleaning their latrines less than once a week; after implementation, this dropped to 5.4% of beneficiary schools on average (Figure 24). Likewise, only 25.3% of beneficiary schools cleaned their latrines twice or more per week at baseline, and on average 61.0% of beneficiary schools cleaned their latrines that frequently after implementation.

As was reported previously, we saw evidence that latrine cleanliness was associated with decreased open defecation. Comparison of reported latrine cleaning frequency and observed latrine cleanliness found that there was no significant difference in observed cleanliness if schools reported cleaning their latrines daily (92.8% of latrines clean) compared to 2-4 times per week (85.9% of latrines clean, p=0.06). However,

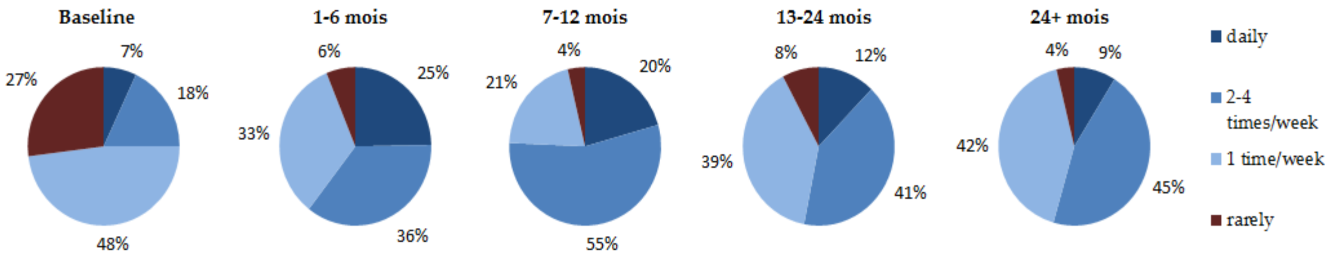
Figure 24: Average reported frequency of latrine cleaning



latrines that were cleaned at 2-4 times per week were more likely to be clean than latrines that were cleaned once per week (71.1% of latrines clean, p<0.01). Cleaning at least once per week was better than cleaning more rarely (45.6% of latrines clean, p=0.01).

The percentage of beneficiary schools that reported cleaning their latrines twice or more per week was highest among schools in the period of 7-12 months after implementation, at 75.7% (Figure 25). However, in the period of 1+ years after implementation fewer beneficiary schools (53.5%) reported cleaning their latrines twice or more per week on average.

Figure 25: Frequency of latrine cleaning at beneficiary schools by time since implementation



D. IMPACTS

Results for impact indicators are presented by the time of year that the data was collected rather than time since program implementation in order to capture seasonal and inter-annual fluctuations.

1. IMPACT ANALYSIS

INDICATORS

One week, period prevalence of absence reported by roll call was our primary outcome. Enumerators used ledgers to collect gender- and grade-identified school-level pupil enrollment, pupil absences, and pupil drop-out due to abandonment, transfer, or death, as verified by the teachers. Current enrollment was calculated by subtracting the count of pupils who had dropped out from the count of total enrollment at each data collection point. Unadjusted mean roll call absence was calculated as the number of absent pupils over the number of currently enrolled pupils. In final models, roll call absence was calculated as a binomial event, where the individual binomial data were aggregated at the school level. Absences that were followed by a designation as abandoned were removed from analysis. Data collected on days where enumerators noted teacher absences, festivals, or other events that could impact attendance were excluded from analysis. Data were also excluded in any class where more than 50% of the pupils were absent, on the assumption that this indicated a special circumstance that was not noted.

Data on the secondary outcomes of self-reported absence, diarrhea, and respiratory infection symptoms in the past week [2] were collected through pupil interviews. All outcomes were binary. Absence was defined as missing a half-day or more. Pupils who reported absence in the last week were asked for the cause or causes of the absence. Pupils who were absent on the day of the survey were included in the calculation of self-reported absence. Pupils were asked if they had had diarrhea using local terminology and were also asked how many times they had defecated each day; a pupil was considered to have had diarrhea if they had defecated three or more times in a day [3]. Pupils were considered to have a respiratory infection symptom if they reported cough, runny nose, stuffy nose, or sore throat. Data for self-reported diarrhea and respiratory infection symptoms were recorded for pupils who were present on the day of the survey only.

ANALYSIS

Data were analyzed using STATA Statistical Software: Release 13 (College Station, TX). Descriptive statistics were calculated by aggregating individual-level data to the school-level where necessary. We used two sample t-tests to assess the differences in means between beneficiary and comparison schools. Unadjusted mean percentages for all outcome variables were calculated by aggregating individual-level data to the school-level.

To quantify the impact of the program, we employed intention-to-treat analyses utilizing mixed-effects logistic regression models that compared beneficiary schools to comparison schools, without regard to program adherence. We only included beneficiary schools and their respective matched schools in the analysis once implementation activities were complete. Models included several design variables, including the educational district (CAP) and the time of year, and random effects to account for clustering of pupils within schools and for repeated measures of pupils over time. We also assessed for effect modifiers and residual confounders for several variables determined a priori, including school size, rural or urban zone, pupil age, pupil sex, pupil grade, and reported latrine presence at home. Covariates were determined to be effect modifiers if an interaction term between the covariate and intervention group was significant in the full model. Covariates were determined as potential confounders if we observed imbalances between potential covariates in the beneficiary and comparison groups. We present fully adjusted models if confounders or effect modifiers were observed.

2. IMPACT RESULTS

DEMOGRAPHICS

Table 8 presents demographic information for the pupils who participated in individual surveys. A total of 4,907 pupils were recruited among beneficiary schools and 4,823 pupils were recruited in the comparison schools. Mean age, mean grade, and household latrine coverage was similar between groups. Baseline data on attendance and health outcomes were not collected, as intervention activities had begun prior to the evaluation period. Due to the general balance of covariates between beneficiary and comparison groups, we did not include any of these covariates in our final models.

Table 8: Descriptive statistics of sample population by study arm for pupils at time of enrollment

	Intervention (*n=4,907*)		Comparison (*n=4,823*)		p†
Age	11.0	(0.8)	10.9	(0.7)	0.22
Grade	4.3	(0.3)	4.2	(0.2)	0.07
Reported household latrine presence	85.1%	(25.5)	82.5%	(29.4)	0.50

Results indicate mean (SD) or mean % (SD)
†P-values based on independent samples t-tests on school-level averages

A total of 887 school visits including 772 roll call measures (393 beneficiary, 379 comparison) and 31,178 pupil interviews (15,681 beneficiary, 15,497 comparison) were conducted after the completion of program activities in beneficiary schools and at corresponding time points for the matched comparison schools

ABSENTEEISM

Roll call absence was higher among pupils in beneficiary schools compared to pupils in comparison schools (Table 9). Rates of absence were comparable between boys and girls and sex was not an effect modifier in the relationship between the intervention and roll call absence. Data was stratified by grade for exploratory analysis. Rates of roll call absence varied between grades, although absence was higher among beneficiary schools across all grades. The relationship was not significant within grade five. Further exploratory analysis looked at grade and sex. The overall relationship between the intervention and increased absence remained consistent, although the study was not powered to detect differences in these sub-populations and few of the relationships were statistically significant.

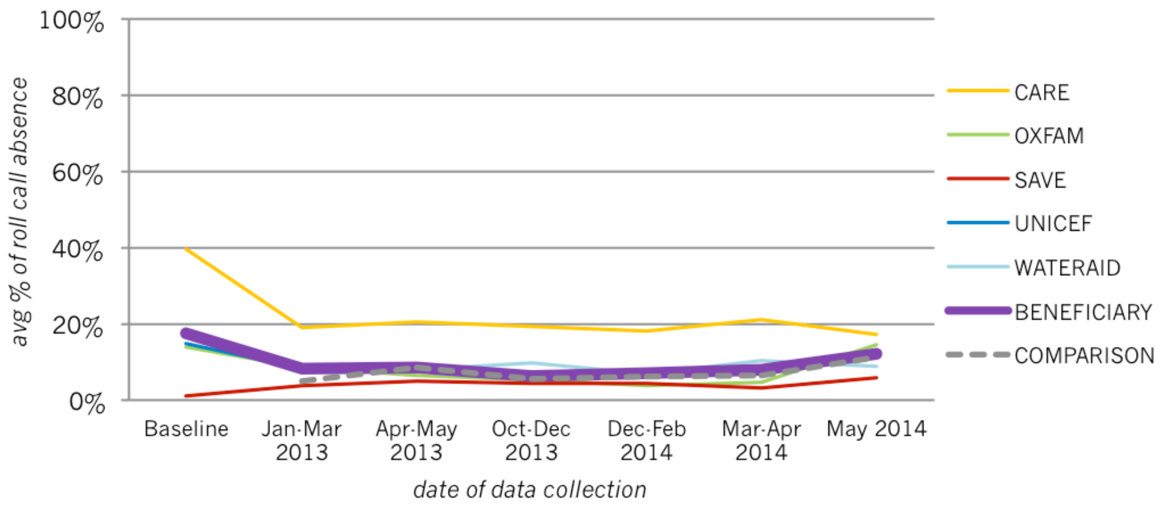
Table 9: Logistic regression models comparing roll call absence between schools that participated in the DCIM WASH intervention and comparison schools

	Intervention *n=100*		Comparison *n=100*		Regression*		
	Mean%	SD	Mean%	SD	OR	95% CI	*p*
Roll call absence, total	8.0	(5.5)	6.7	(5.6)	1.25	1.08,1.44	**<0.01**
Stratified by sex							
Girls	8.0	(6.0)	6.6	(5.2)	1.24	1.07,1.44	**<0.01**
Boys	8.1	(5.2)	6.8	(5.8)	1.24	1.07,1.44	**<0.01**
Stratified by grade							
Grade 1	8.8	(6.0)	7.4	(7.6)	1.22	1.00,1.48	**0.05**
Grade 2	8.3	(6.6)	6.9	(6.5)	1.28	1.06,1.55	**0.01**
Grade 3	6.8	(5.7)	5.9	(6.1)	1.19	0.99,1.43	**0.01**
Grade 4	7.7	(6.3)	5.8	(5.4)	1.30	1.08,1.56	**<0.01**
Grade 5	7.3	(6.3)	6.7	(6.1)	1.07	0.89,1.29	0.48
Grade 6	8.3	(6.3)	6.8	(6.6)	1.27	1.03,1.55	**0.02**
Stratified by grade and sex							
Girls Grade 1	8.9	(7.0)	7.8	(8.2)	1.21	0.99,1.48	0.06
Girls Grade 2	8.5	(7.9)	6.9	(7.7)	1.35	1.07,1.70	**0.01**
Girls Grade 3	6.8	(6.1)	5.3	(6.4)	1.34	1.09,1.67	**<0.01**
Girls Grade 4	7.6	(7.3)	6.3	(6.5)	1.17	0.95,1.43	0.14
Girls Grade 5	7.1	(6.6)	5.6	(6.0)	1.15	0.93,1.44	0.20
Girls Grade 6	7.6	(7.1)	6.2	(6.9)	1.25	0.99,1.57	0.06
Boys Grade 1	8.8	(6.6)	6.9	(6.6)	1.24	0.99,1.55	0.06
Boys Grade 2	7.9	(6.6)	6.6	(6.2)	1.21	0.99-1.47	0.07
Boys Grade 3	6.5	(5.8)	6.4	(6.7)	1.06	0.87,1.29	0.58
Boys Grade 4	7.8	(6.5)	5.4	(5.2)	1.38	1.13,1.69	**<0.01**
Boys Grade 5	7.4	(6.8)	7.1	(7.0)	1.00	0.80,1.23	0.96
Boys Grade 6	8.7	(6.9)	7.2	(7.5)	1.26	0.99,1.59	0.06

* All models included variables for the intervention, the matched cluster, and the time of year, and accounted for clustering of pupils within schools and for repeated measures of pupils over time.

Figure 26 presents roll call absence data by time that the data was collected. On average, roll call absence was lower among beneficiary schools after the intervention than it was at baseline (17.5% baseline, 8.0% post-intervention). Absence was the highest among CARE schools and lowest among Save the Children schools. Rates of absence were similar across other partner schools. There is some indication of a seasonal trend of increased absence in the April-May period, which corresponds to the end of the school year and school exams.

Figure 26: Average roll call absence by time of data collection



We did not find any difference in total self-reported absence between beneficiary and comparison schools (Table 10). Pupils in beneficiary schools who reported absence in the past week were less likely to report absence due to diarrhea than pupils in comparison schools. There was no difference in reported absence due to respiratory infection symptoms. Pupils in the interview sample were more likely to be absent in beneficiary schools compared to comparison schools. Pupils in the interview samples in beneficiary schools were more likely to be absent on the day of the survey than pupils in the comparison schools. Rates of absence from this sub-sample were comparable to roll-call absence rates.

Table 10: Logistic regression models comparing pupil-reported absence in the past week between schools that participated in the DCIM WASH intervention and comparison schools

	Intervention *n=4907*		Comparison *n=4823*		Regression*		
	Mean%	SD	Mean%	SD	OR	95% CI	*p*
7-day absence recall, total	17.6	(8.8)	18.0	(8.9)	0.93	0.79,1.09	0.38
7-day absence recall, girls	18.2	(10.6)	19.0	(10.0)	0.86	0.72,1.04	0.12
7-day absence recall, boys	17.0	(8.5)	17.1	(9.7)	1.00	0.82,1.20	0.97
% of absences attributed to diarrhea	24.1	(17.3)	27.4	(16.1)	0.73	0.56,0.94	**0.02**
% of absences attributed to respiratory infections	8.2	(9.2)	8.2	(9.3)	0.96	0.67,1.38	0.83
% of pupils in sample absent on day of survey	7.6	(6.5)	6.3	(6.7)	1.26	1.02,1.55	**0.03**

* All models included variables for the intervention, the matched cluster, and the time of year, and accounted for clustering of pupils within schools and for repeated measures of pupils over time.

We collected data on school enrollment, abandon rates, and passing rates in order to assess additional educational impacts from the DCIM program. However, due to external factors related to the 2011 coup and 2013 military actions, there was a high degree of enrollment pressure on the evaluation schools; 66% of beneficiary schools and 54% of comparison schools reported exceptional increases or decreases in enrollment during the evaluation period. These disruptions made our ability to detect an association between enrollment and passing rates and the DCIM program unreliable. We therefore have not included this data in the final analysis.

DIARRHEA

Pupils in the beneficiary group were less likely to report diarrhea in the past week than were pupils in the comparison group (Table 11). We found that school size was a potential effect modifier. In a stratified analysis, this effect was primarily seen among schools in the bottom two quartiles of population size. As these data were collected only for pupils who were present on the day of the survey, we ran a sensitivity analysis probabilistically imputing diarrhea to pupils who were not present. Imputing diarrhea to 36.5% of the absent pupils (the overall proportion of present pupils who reported an absence in the last week due to diarrhea) we continued to see an effect (OR 0.78, 95% CI 0.67-0.91, p<0.01).

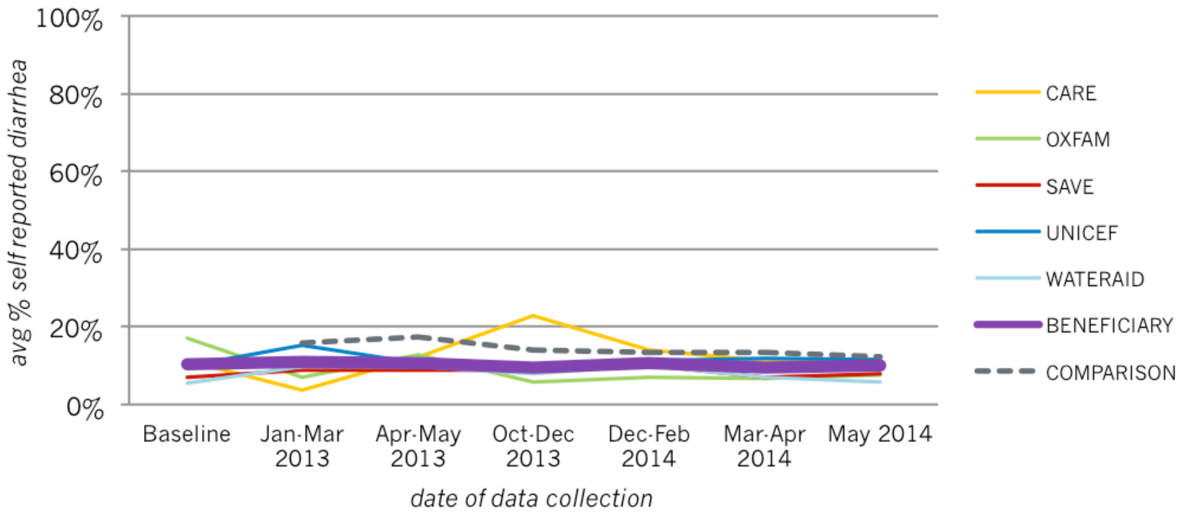
Table 11: Logistic regression models comparing pupil-reported diarrhea in the past week between schools that participated in the DCIM WASH intervention and comparison schools

	Intervention *n=4907*		Comparison *n=4823*		Regression*		
	Mean%	SD	Mean%	SD	OR	95% CI	*p*
7-day diarrhea recall, total	10.0	(4.7)	13.0	(7.9)	0.71	0.60,0.85	**<0.01**
7-day diarrhea recall, girls	10.2	(6.3)	13.0	(8.1)	0.72	0.57,0.09	**<0.01**
7-day diarrhea recall, boys	9.8	(6.3)	13.0	(8.1)	0.68	0.56,0.82	**<0.01**
Stratified by school size							
Quartile 1 (31-158 pupils)	10.0	(3.1)	18.0	(10.5)	0.43	0.28,0.66	**<0.01**
Quartile 2 (159-299 pupils)	9.9	(6.0)	13.0	(5.4)	0.69	0.49,0.97	**0.03**
Quartile 3 (300-441 pupils)	10.2	(4.3)	8.9	(4.8)	0.92	0.72,1.17	0.47
Quartile 4 (442-1129 pupils)	9.9	(5.0)	10.7	(4.9)	0.79	0.57,1.09	0.15

* All models included variables for the intervention, the matched cluster, and the time of year, and accounted for clustering of pupils within schools and for repeated measures of pupils over time.

Figure 27 presents data for self-reported diarrhea incidence in the past week by time that the data was collected. We did not see any seasonal trends in diarrheal self-report. Rates remained fairly constant across all time points among beneficiary schools and were consistent across partners. Pupils in CARE schools showed the greatest degree of variability, with elevated diarrhea prevalence in the October-December period. Rates of self-reported diarrheas were consistently higher in comparison schools compared to beneficiary schools, although rates appeared to decline over time.

Figure 27: Average one-week diarrhea self-report by time of data collection



RESPIRATORY INFECTION

Pupils in the beneficiary group were less likely to report having had at least one symptom of respiratory infection (cough, runny nose, stuffy nose, or sore throat) than were pupils in the comparison group (Table 12). We ran a sensitivity analysis imputing respiratory infection symptoms to pupils who were not present on the day of the survey. Imputing respiratory infection symptoms to 55.3% of the absent pupils (the rate of respiratory infection incidence for all present pupils plus the percentage of absences reported as due to respiratory infections) we continued to see an effect (OR 0.76, 95% CI 0.67-0.87, p<0.01).

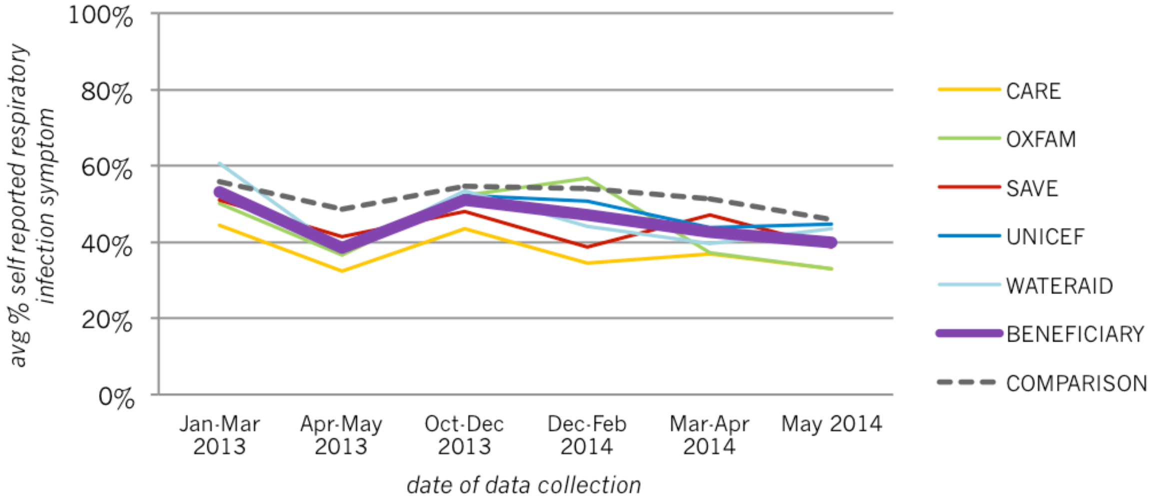
Table 12: Logistic regression models comparing pupil-reported respiratory infection (RI) symptoms in the past week between schools that participated in the DCIM WASH intervention and comparison schools

	Intervention *n=4907*		Comparison *n=4823*		Regression*		
	Mean%	SD	Mean%	SD	OR	95% CI	*p*
7-day RI symptom recall, total	45.2	(12.2)	51.2	(13.0)	0.75	0.65,0.86	**<0.01**
7-day RI symptom recall, girls	45.2	(14.4)	51.3	(15.1)	0.74	0.62,0.87	**<0.01**
7-day RI symptom recall, boys	45.6	(12.8)	51.2	(12.8)	0.76	0.65,0.89	**<0.01**

* All models included variables for the intervention, the matched cluster, and the time of year, and accounted for clustering of pupils within schools and for repeated measures of pupils over time.

Figure 28 presents data for self-reported respiratory infection symptom incidence in the past week by time that the data was collected. There is some indication of seasonal patterns in report of respiratory infection, with increased rates in October-January (the cold season) and lower rates in April-May (hot, dry season). Respiratory infection was lowest among CARE schools in Mopti region, and similar across the other evaluation areas.

Figure 28: Average one-week respiratory infection self-report by time of data collection



E. DISCUSSION

This evaluation quantified the impact of the Dubai Cares WASH in Schools Initiative in Mali – one of the largest and most comprehensive school WASH programs undertaken to date – on pupil health and education. Beneficiary schools were found to meet all 15 WASH criteria that the partners had set as outcome goals at only 25% of visits on average; however, the program schools exhibited considerable gains in WASH access and behaviors among pupils compared both to the comparison schools and to the program schools at baseline. Beneficiary schools met 10 or more of the 15 WASH criteria at 87% of visits. Beneficiary schools sustained most of these gains over time. We found evidence of reductions in self-reported diarrheal incidence and incidence of respiratory infection symptoms among pupils attending beneficiary schools compared too those in the comparison schools. We found no impact of the program on reductions in pupil absence, and saw higher rates of roll call absence among beneficiary schools compared to comparison schools. Higher roll call absence rates seen in beneficiary schools may have been due to the intervention, but are more likely the result of an imbalance in unobserved confounders between study groups at baseline.

OUTPUT AND OUTCOMES

We collected rigorous outcome data to inform program partners on their success in ensuring that key program targets were met and sustained in the schools where they worked. In order for a healthy WASH environment to be sustained in schools, there must not only be sufficient infrastructure and supplies, but also sufficient governance at the school level to ensure that infrastructure is maintained, supplies are replaced, and appropriate behaviors are encouraged. Overall, schools benefiting from the DCIM program had improved school WASH environments at compared to the situation of those schools at baseline and the matched comparison schools.

WASH STANDARDS

Beneficiary schools performed much better at meeting the 15 WASH criteria set by the DCIM partners, meeting 13 or more criteria at half of all visits, compared to comparison schools where fewer that 6 criteria were met at half of the visits. Among the four WASH standards that the partners established, the water access standard was most frequently attained. Where environmental conditions and funding allowed, the DCIM partners ensured that improved water points were available on all school grounds. The greatest barrier to meeting this standard was the functionality of the water points (met at 83.0% of visits), and this rate remained constant over time. This indicates that there were some gaps in repair and maintenance of water points, but that the rate of repair was sustained even after two years following implementation. Water points that were shared with the community were equally likely to be functional as water points that were not shared, which provides little evidence on the role of communities in either supporting or misusing the infrastructure. Achieving the water access standard was a challenge for CARE due to environmental and financial constraints. However, despite the lack of water points on all school grounds, CARE schools ensured that pupils had access to drinking water 83.3% of the time.

Beneficiary school also fared well at sustaining hygiene supplies; the hygiene kit standard was met at 74.0% of observations in beneficiary schools, compared to 11.7% at baseline and 7.3% among comparison schools. Beneficiary schools were particularly successful at maintaining durable supplies such as water containers and anal cleansing kettles, which were available from 94.1-97.9% of the time on average and were consistently present over time. Consumable supplies such as soap, detergent, and bleach were less likely to be present, and were available from 70.9-81.6% of the time on average. Detergent and bleach were less likely to be present at schools as time since implementation increased, although the percentage of schools were soap was available remained constant over time. This indicates that repurchase of soap may have been prioritized over purchase of other consumable materials among beneficiary schools. These indicators measured the presence of "at least one" of each element of the hygiene kit, rather than the absolute number of each supply that was available. There may have been an unseen decrease over time in the number of supplies that were available.

Schools faced greater challenges in meeting the handwashing standard, which was met in beneficiary schools at 57.9% of observations. Although handwashing containers were present in schools at 89.8% of observations, schools did not always ensure that these containers had water inside (observed 73.7% of visits) and soap nearby (observed at 57.9% of visits). Ensuring that pupils had access to soap was a particular challenge; although handwashing soap was observed to be present at beneficiary schools 81.6% of the time, soap was observed to be available to pupils at 69.2% of observations, and specifically made available next to a handwashing station with water 57.9% of the time. This is higher than what was found in a similar trial in Kenya, where soap was found in only 40% of schools, though in Kenya it was not provided directly by the program[4].

The discrepancy between soap being present at schools and being available to pupils may in part be due to the practice of reserving soap in the school director's office in order to prevent loss or theft, which discourages pupils from using soap for handwashing. Schools were slightly less likely to achieve the handwashing standard as time since implementation increased, with 67.4% meeting the standard 1-6 months after implementation and 57.3% at 2+ years after implementation. However, this decrease was slight and indicates that hadwashing containers with soap and water were made available to pupils consistently over time.

On average, cement handwashing containers were observed to have water and functional taps only 22.7% of the time. This figure declined from a peak of 44.8% among schools 1-6 months after implementation to 0.0% among schools 2+ years after implementation. While permanent cement handwashing stations that are located near the latrines have an appealing potential to enhance handwashing practice after latrine use, in reality they have been rarely utilized. Investigations into the barriers that prevent their use in this context are warranted.

Among the four WASH standards set by the partners, the greatest challenge for most schools was in meeting the sanitation standard. Nearly all beneficiary schools were observed to have latrines available to pupils on school grounds. Overall, beneficiary schools achieved a ratio of 60 pupils per latrine, well below the benchmark of 70 pupils per latrine that was set by partners. However, only 72.9% of schools met the threshold of 70 pupils per latrine. This was in part due to the practice of locking latrine doors, which some schools undertake to prevent misuse or to alternate which latrines are used. Latrines were not included in calculation of the standard if they were locked; even if keys are available to pupils on demand, this practice can prevent the use of latrines. Challenges in meeting the pupil:latrine ratio may have also been due in part to changing enrollment in the schools. Schools also faced challenges in enforcing separation of the latrines by sex (76.8% observed), and in ensuring that the latrines were clean (1 clean latrines per 70 pupils observed at 56.6% of visits). Overall, schools were less likely to meet the sex-separation, safety, and cleanliness criteria within the sanitation standard as time since implementation increased, indicating potential problems with sustainability. The latrine standard included a number of criteria that were dependent on action at the schools level, including cleaning latrines, enforcing sex separation, and repairing damages. Low performance in this area highlights the difficulty of ensuring ownership and oversight of day-to-day WASH maintenance among school stakeholders.

There were a few key differences in achievement of WASH standards between urban and rural schools. Urban schools were more likely than rural schools to meet the water access standard, possibly due to increased access to piped water systems in urban areas. Urban schools were also more likely to enforce sex separation of latrines. Rural schools, on the other hand, were more likely to meet the overall handwashing standard and specifically more likely to ensure availability of water for handwashing; this association may also be due to the fact that rural schools tended to be smaller and such facilities may have been easier to monitor.

Larger schools had a more difficult time meeting overall latrine ratio and latrine cleanliness criteria than smaller schools. This may be a factor of changing enrollment; as school population increases, schools are less likely to achieve the preset ratio with their existing latrines. It may also have been due to the challenge of monitoring numerous latrines and the pressure that larger numbers of students bring to the facilities. School size did not appear to be a factor in achievement of any of the other standards or criteria.

BEHAVIOR CHANGE

The program was effective in achieving behavior change on the part of the pupils. Signs or report of open defecation were recorded at only 23.8% of schools on average during the evaluation period, compared to 61.9% at baseline and 75.3% of comparison schools. This rate was sustained as time since implementation increased, and schools where implementation had been completed more than two years prior were slightly more likely to be free of open defecation (85.4%) than schools where implementation had been completed 1-6 month prior (79.8%). This improvement may indicate that the behavior takes some time to adopt, although it could also be a factor of different behavior change promotion strategies used in the different sets of schools.

Rates of handwashing with soap in beneficiary schools increased from 35.2% at baseline to 56.7% during the evaluation period. Rates of handwashing with soap varied over time, with peak rates observed among schools 6-12 months following implementation of the DCIM program. As with open defecation, the peak seen in the 6-12 month period may indicate that behaviors take time to adopt, or may be a factor of a differing implementation approaches in different batches of schools.

GOVERNANCE

Governance of school WASH infrastructure and hygiene promotion programs is seen as a key driver of sustainability. Beneficiary schools performed better at all governance indicators compared to baseline and comparison schools including designated a teacher, pupil or parent to be responsible for WASH, spending money on WASH, cleaning latrines twice a week or more, and teaching WASH lessons to all classes. The areas where beneficiary schools had the most room for improvement were in having a participatory and active SMC (43.5% - especially participation of pupils was an issue) and having designated a budget for WASH (36.3%); however, beneficiary schools were still more likely to meet these indicators than were comparison schools or at baseline.

Governance indicators were associated with several WASH outcomes. These associations do not necessarily mean that meeting the governance indicator *caused* improved or worsened performance, but they may highlight areas for further investigation. Schools with SMCs that met regularly and included a parent and a pupil were less likely to meet the water access standard, but there was no association between any of the other DCIM standards. This negative relationship may serve as an indication that the frequency of and attendance at SMC meetings are not sufficient metrics for determining the efficacy of a school's governance system.

Schools that reported having a budget in place for WASH were more likely to meet the water access, handwashing, and hygiene kit standards, and schools that reported having spent money on WASH were more likely to meet the water access standard. These associations may reflect the direct effect of increased investment in WASH infrastructure, or the association could be related to external factors such as increased overall prioritization of WASH by school stakeholders. The spending patterns between intervention and comparison schools suggest that, by providing infrastructure and consumables, the DCIM program allowed beneficiary schools to invest more funds in the long-term upkeep of their WASH infrastructure since schools no longer needed to invest in the facilities and supplies provided. It may also indicate an increased awareness among beneficiary schools of the need for investing in maintenance.

Beneficiary schools with a teacher, parent or pupil designated as responsible for WASH were more likely to meet the sanitation standard. This association could be a direct result of increased oversight of the latrines, or it could be that both factors are associated with increased prioritization of WASH by school stakeholders. There was an association between an increased percentage of clean latrines and decreased open defecation, and latrines were more likely to be clean if they were cleaned twice or more per week. Although beneficiary schools cleaned their latrines more frequently than comparison schools, having a sufficient number of clean latrines was the criteria within the WASH standards that was most difficult overall for beneficiary schools to meet, suggesting that implementing partners could strengthen their promotion of cleaning practices.

Beneficiary schools did not meet all of the standards set by DCIM partners and did not fully maintain the WASH infrastructure and supplies given to them by program partners. There was also some indication of challenges to sustainability, particularly in replacing consumable supplies, ensuring that handwashing containers with soap and water were available to pupils, and ensuring that there were a sufficient number of safe, clean, sex-separated latrines available to pupils. However, beneficiary schools performed better in all output and outcome indicators tracked by this evaluation, even over two years after implementation. Even with declines over time, beneficiary schools remained well above baseline levels in terms of the WASH environment or pupil handwashing and open defecation behaviors.

IMPACTS

We did not find a significant difference in the rates of self-reported absence between pupils attending beneficiary versus comparison schools, and we found that pupils in beneficiary schools were 23% more likely to be absent at roll call than were pupils in the matched comparison schools. This contrasts with previous studies that found lower rates of absenteeism after the implementation of school WASH programs [4-6]. The fact that we found no effect modification based on sex rules out differential impacts among girls, as found in Kenya [7], nor did we see a relative improvement in absence rates in beneficiary schools among older girls who might be expected to benefit from improved ability to manage menstruation based on improved access to WASH facilities.

Although it is possible that the increased levels of roll call absenteeism observed in beneficiary schools was due to the program, a more likely explanation is that program schools had higher rates of absence than comparison schools at the outset of the study, and our matching criteria was insufficient to control for unobserved confounders. This is one of the risks associated with a non-randomized study design. The loss to follow up of schools that participated in the baseline assessment largely due to the ongoing conflict meant that we did not have baseline absenteeism and health measures for the majority of schools in the study, which could have allowed a stronger assessment of comparability between the beneficiary and comparison groups. However, we were able to match on several important school characteristics, and we observed general balance of covariates between study groups.

Another possible explanation for higher rates of absence among beneficiary schools is the potential for the program itself to have increased exposure to fecal pathogens among pupils. There is some evidence from the literature that improvements to WASH infrastructure and conditions may lead to increased exposure to fecal pathogens. Greene et al found that pupils in schools that received latrines had higher risk of having *E. coli* on their hands [8], and the authors hypothesized that this may have been due to increased use of latrines that was not accompanied by commensurate increases in handwashing behavior. Although we saw that availability of soap and clean latrines was improved at beneficiary schools compared to comparison schools, we also saw that soap was not available to pupils at 30% of observations in beneficiary schools, that even when soap was available less than 60% of pupils washed their hands with soap upon leaving the latrines, and that beneficiary schools provided one clean latrine per 70 pupils at less than 60% of observations. These gaps in coverage of handwashing materials, handwashing behavior, and access to clean latrines may have led to increases in exposure as pupils used their new sanitation facilities. Pupils did report lower rates of diarrhea and respiratory infections in beneficiary schools, and absence due to illness from diarrhea, indicating that this hypothesis may not be applicable in this setting; however the potential for a negative relationship between school WASH improvements and absenteeism may be worth exploring.

In contrast to the absenteeism findings, we found evidence that a comprehensive school-based water, sanitation, and hygiene intervention can have a positive impact on pupil health. Pupil in the beneficiary group were 29% less likely to report having had diarrhea in the past week compared to pupils in comparison schools, and were 25% less likely to report having had symptoms of respiratory infections. Our findings are comparable with research showing that school WASH interventions are associated with lower rates of diarrhea self-report and influenza incidence [9, 10] and help strengthen the evidence that school WASH interventions may reduce exposure pathways for diarrheal diseases and promote healthy WASH behaviors among pupils. In addition, we found that pupils who had been absent in the last week in beneficiary schools were 27% less likely to report their absence being due to diarrhea. These data are comparable with findings of reduced absence due to diarrhea and illness after a handwashing intervention [10]. However, the reductions in absence due to diarrhea may not have been great enough to overcome any pre-existing imbalances in absence rates between the beneficiary and comparison schools.

STRENGTHS AND LIMITATIONS

There were several aspects of the evaluation that limit the internal and external validity of our findings. The first and most substantial limitation was our study design. We employed a matched control design, a common non-experimental approach, due to the preference of Dubai Cares to intervene on larger schools and due to the timeframe in which we were engaged in the study. Matching criteria included location, which was intended to serve as a proxy for both environmental conditions and community/household characteristics; school size; and WASH status. We were successful at matching on location and WASH status, but less successful at matching on school size, given that many of the largest schools had been selected as beneficiaries in order to maximize cost-effectiveness of the program. Matching allows for control of observed characteristics, but there is no way to control for unobserved characteristics. A randomized control trial (RCT) is an experimental design considered the gold standard of causal evidence where clusters (in this case, schools) are randomly allocated to intervention and control status prior to baseline. By contrast with a matched design, this approach controls for *all* observed and *unobserved* factors. In the case of this evaluation, the selection process for counterfactual (comparison schools) may not have been sufficiently rigorous to eliminate potential bias.

A second limitation of our evaluation design was that we did not have baseline data for all schools. Due to the combined effect of armed conflict in the North of the country and the anticipated implementation schedule not being met, half of the schools that participated in the baseline study were not eligible for the impact evaluation. While our selection of replacement schools was rigorously executed, many of those replaced schools did not have a true baseline as program implementation was conducted before data collection could take place. Typically, baseline data is not necessary as part of a longitudinal surveillance to control for the difference between intervention and comparison schools, though it is useful in order to show that balance exists between the two groups.

A third limitation was that our primary illness measures were based on self-report by pupils. Self-report measures are known to be subject to bias. Although there is no evidence to suggest that any bias may have existed to a greater degree among the beneficiary group compared to the comparison group, our study was not blinded for either the beneficiaries or data collectors. We therefore have no way to determine if bias was responsible for some or all of the difference we detected between these groups for self-reported diarrhea and respiratory infection.

A fourth limitation is related to our inability to report on the sustainability of the intervention in absolute terms. While we present our outcome data in terms of time since intervention, we are not able to characterize that as the sustainability of the intervention. This is because we were not able to collect data in all schools at the same time points after implementation; for some schools, our data collection started 1-2 years after the intervention had been completed ("older" schools), while for others we only tracked schools for 1-6 months after implementation ("newer" schools). Implementation approaches also changed over time. We cannot say whether the newer schools will perform the same as the older schools in the future as a result of changes to implementation strategies. In order to have an accurate measure of sustainability, we ideally would need to follow a single cohort of schools over the same time period.

While there were several limitations of this evaluation discussed above, the overall methods and approach used were quite robust. This is one of the largest and most comprehensive impact evaluations to date of a school WASH program in low-income settings. We utilized a strong matched design and followed schools over time in order to account for seasonal variations. Our primary impact measure, absence, was collected using roll-call data collected by our surveillance team, which is a highly objective measure of school absence. In addition, we assessed a broad range of outcome-based indicators, which will allow for additional analysis to assess the effect of WASH conditions on the impact indicators.

This impact evaluation was conducted by external evaluators from Emory University, ensuring unbiased assessment of the program. Evaluators worked closely with the implementing partners throughout the program to ensure that outcome data could iteratively inform program learning. The evaluation of several programmatic approaches from several partners in various contexts within Mali allowed for considerable learning both within and between partners.

F. CONCLUSION AND RECOMMENDATIONS

This evaluation provided evidence that the DCIM intervention improved the WASH environment in beneficiary schools, and that those improvements were sustained two years after implementation of the program. We found that the program produced changes in pupils' handwashing and open defecation practices. We saw evidence that the program had a positive effect on pupils' self-reported incidence if diarrhea and respiratory infection symptoms. However, we did not see evidence of an effect on self-reported absence in the past week, and pupils in beneficiary school were more likely than pupils in comparison schools to be absent according to roll call data. Future analyses will assess the role of the observed WASH environment on pupils' health and absenteeism, as well as accounting for baseline levels of diarrhea self-report and absenteeism.

Based on our experience with this evaluation, we propose several recommendations:

1. This evaluation identified key bottlenecks to achievement of a suitable WASH environment. These included repair of water points, cleaning of latrines at least twice a week and enforcing sex separation in latrines as key factors for ending open defecation practices, ensuring that handwashing containers were filled with water (and avoiding cement handwashing stations), making soap available to pupils, and repurchasing consumable supplies such as soap and detergent. These factors all require an emphasis on capacity building among school stakeholders as well as incentives mechanisms so that they have the ability and willingness required to ensure the correct operation and maintenance of WASH facilities.

2. For a true assessment of the sustainability of the program, the same set of schools would need to be followed for a longer period of time. We were only able to observe 35 schools at a time point of 2 years following implementation of the program, and those schools had not benefitted from the changes that the implementing partners made in the program implementation strategy. Further follow-up of other schools in the evaluation sample could provide a better understanding of how well output and outcome targets were sustained over time.

3. This evaluation documented the state of the school WASH systems, but it did not collect data on the factors that determine which schools fared well and which fared poorly. While we did collect data on several governance indicators, we saw limited associations between these indicators and observed WASH outputs and outcomes at the schools. Further investigation is needed to determine which elements of the governance system play the greatest role. A comparison of schools that did and did not succeed in maintaining suitable WASH environments would help uncover critical factors for sustained success achieving WASH standards and would provide considerable value to the sector and for ongoing WASH in schools programs in Mali. Future programs could focus on those factors in their software strategies, and monitoring programs should incorporate evidence-based indicators to ensure that they are tracking the elements that are most likely to influence WASH outcomes.

4. Close collaboration between the external evaluators and internal monitoring and evaluation staff allowed for increased utility of the data being collected, so that it informed both an assessment of the impact of the program as well as providing monitoring data. However, the evaluators were not brought fully onboard until after baseline data collection which provided some challenges with study design. Engagement of external evaluators at the outset of the program adds

considerable value to the research and learning of a program and can rigor and external validation of the evaluation.

5. The match-control design was sub-optimal. Randomized allocation of the program to schools was not possible in this evaluation due to the desired targeting approach of the program funder. Future studies of this magnitude should consider a randomized design to ensure minimal bias between intervention and comparison groups.

REFERENCES

1. Arnold, B.F., et al., *Simulation methods to estimate design power: an overview for applied research.* BMC medical research methodology, 2011. **11**(1): p. 94.

2. Arnold, B.F., et al., *Optimal recall period for caregiver-reported illness in risk factor and intervention studies: A multicountry study.* American journal of epidemiology, 2013. **177**(4): p. 361-370.

3. Baqui, A.H., et al., *Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes.* International journal of epidemiology, 1991. **20**(4): p. 1057-1063.

4. Freeman, M.C., et al., *Assessing the impact of a school-based water treatment, hygiene and sanitation programme on pupil absence in Nyanza Province, Kenya: a cluster-randomized trial.* Trop Med Int Health, 2012. **17**(3): p. 380-91.

5. Bowen, A., et al., *A cluster-randomized controlled trial evaluating the effect of a handwashing-promotion program in Chinese primary schools.* Am J Trop Med Hyg, 2007. **76**(6): p. 1166-1173.

6. O'Reilly, C., et al., *The impact of a school-based safe water and hygiene programme on knowledge and practices of students and their parents: Nyanza Province, western Kenya, 2006.* Epidemiol Infect, 2008. **136**: p. 80-91.

7. Freeman, M.C., et al., *Assessing the impact of a school-based water treatment, hygiene, and sanitation program on pupil absence in Nyanza Province, Kenya: A cluster-randomized trial.* Trop Med Int Health, 2012. **17**(3): p. 380-391.

8. Greene, L.E., et al., *Impact of a school-based hygiene promotion and sanitation intervention on pupil hand contamination in Western Kenya: a cluster randomized trial.* The American journal of tropical medicine and hygiene, 2012. **87**(3): p. 385-393.

9. Freeman, M.C., et al., *The impact of a school-based water supply and treatment, hygiene, and sanitation programme on pupil diarrhoea: a cluster-randomized trial.* Epidemiol Infect, 2013. **142**(2): p. 340-51.

10. Talaat, M., et al., *Effects of hand hygiene campaigns on incidence of laboratory-confirmed influenza and absenteeism in schoolchildren, Cairo, Egypt.* Emerg Infect Dis, 2011. **17**(4): p. 619-25.