ePact
Strengthening evaluation effectiveness and impact

Monitoring, Evaluation, and Verification Component of the WASH Results Programme

Evaluation Synthesis Report - Volume 2.3 – SAWRP Case Study

Oxford Policy Management

itad

e-Pact is a consortium led by Oxford Policy Management and co-managed with Itad in association with the Institute for Fiscal Studies.

This evaluation was commissioned by the UK Department for International Development and was funded with UK aid from the UK government.

UKaid
from the British people

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies.

Acknowledgements

This case study report is the culmination of evaluation activities undertaken by Oxford Policy Management (OPM) between November 2014 and January 2019.

We would like to thank Guy Howard, Stephen Lindley-Jones, Leonard Tedd, and Laura Westcott at the UK Department for International Development, the suppliers and their partner, and the monitoring and verification team for their time and inputs during this period.

This case study was written by Zach White and Jeremy Colin with support from Sarah Javeed. The case study was reviewed internally by Timothy Forster, Julia Larkin and Lucrezia Tincani.

The evaluation team are extremely thankful to all those on SAWRP who facilitated the fieldwork. In particular: John Dean, Teia Rogers, and Abdul Saboor.

Table of contents

List of tables, figures, and boxes

Abbreviations

ASWA	Accelerating Sanitation and Water for All
AWARE	Association for Water, Applied Education, and Renewable Energy
BCC	Behaviour change communication
CCU	Consortium Coordination Unit
CLTS	Community-led total sanitation
CRP	Community resource person
DEQ	Detailed evaluation question
DFID	UK Department for International Development
FSM	Faecal sludge management
HEQ	High-level evaluation question
INGO	International non-government organisation
IP	Implementing partner
JMP	Joint Monitoring Programme for Water and Sanitation
KPK	Khyber Pakhtunkhwa
LGD	Local Government Department
LPP	Lodhran Pilot Project
MDG	Millennium Development Goal
MSAN	Multi-Sector Action on Nutrition
MV	Monitoring and Verification
MVE	Monitoring, Verification, and Evaluation
MVOC	Monitoring and Verification Framework for Outcomes
MVOs	Monitoring and Verification Frameworks for Project Outputs
NGO	Non-governmental organisation
NOC	No objection certificate
NRSP	National Rural Support Programme
ODF	Open defecation free
OPM	Oxford Policy Management
PATS	Pakistan Approaches to Total Sanitation
PbR	Payment by results
PHED	Public Health Engineering Department
RAG	Red/amber/green
RCT	Randomised control trial
RuSFAD	Rural Sanitation in Flood-Affected Districts
SAF	Sustainability Assessment Framework
SAWRP	South Asia WASH Results Programme
SDG	Sustainable Development Goal
SSH4A	Sustainable Sanitation and Hygiene for All
SWIFT	Sustainable WASH in Fragile Contexts
TOC	Theory of change
UC	Union Council
UNICEF	UN Children's Fund
WASH	Water, sanitation, and hygiene
WEDC	Water, Engineering and Development Centre at Loughborough University
WHO	World Health Organization
WSUP	Water and Sanitation for the Urban Poor

1 Introduction

1.1 Structure of the WRP reporting

The endline evaluation of the WRP is presented across five separate volumes. This is due to the size and complexity of the programme. Vol. 1 presents the summary findings across the three Suppliers and addresses the evaluation questions. Vol. 1.2 contains the annexes to the summary report. Vols. 2.1, 2.2, and 2.3 are supplier-specific case studies and seek to provide far greater contextual information and discuss the evidence under the relevant thematic headings.

Table 1: Evaluation findings reporting

Volume		Contents
Vol. 1.1	Summary report	- These volumes address the evaluation questions across the three suppliers. - Provides lessons and recommendations
Vol. 1.2	Summary report annexes	
Vol. 2.1	SWIFT case study	*See below for details of the report structure* *The structure of the three case studies is the same to allow for issues to be compared across the three suppliers.*
Vol. 2.2	SSH4A case study	
Vol. 2.3	SAWRP case study	

1.2 Purpose of this report

This report summarises the main findings of the WASH Results Programme endline evaluation, for the SAWRP programme only. It summarises the findings from the country visit to Pakistan in early 2018; remote interviews with SAWRP programme managers in the UK and Bangladesh; a review of relevant documentation, and incorporates finding from the midline assessment where relevant. In addition the report draws on insights gained from round table meetings and learning events with suppliers and verifiers during the output and outcome phases of the programme.

This report does not seek to answer the evaluation questions directly as these pertain to the programme as a whole, and are addressed directly in Vol. 1.1. Rather, this report notes the evaluation team's reflections as they pertain to SAWRP across the relevant thematic areas. This report is best viewed as documenting the supporting evidence from SAWRP that contributes to addressing the evaluation questions.

1.3 Structure of the remainder of this report

- **Section 2** provides details of the evaluation approach with reference to the SWIFT specific data collection, limitations, and potential sources of bias;
- **Section 3** provides an overview the SWIFT consortium, the results, and the implementation in each of the SWIFT counties; and
- **Section 4** discusses the findings of the evaluation under each of the DAC criteria and under thematic headings related to the evaluation questions.

2 Evaluation approach for SAWRP

2.1 Stakeholders met

Evaluation activities	Pakistan	Bangladesh	Comment (e.g. if not met / only able to meet partially)
Supplier global managers	✓	✓	Plan International, WaterAid only at endline.
Lead verifier	✓		
Country verifiers	✓		
Learning partner	✓		Interviewed at endline only
Supplier country staff	✓	✓	Plan International, WaterAid only at endline, and Unilever in Pakistan at midline. Remote interviews conducted with Bangladesh country management (Plan and WaterAid) staff at endline
Local implementing partners (IPs)	✓	×	All IPs interviewed in Pakistan at midline and endline
Government counterparts	✓	×	Provincial and local government in Pakistan
Community members (service users)	×	×	Not part of the planned evaluation activities
Other sector experts*	✓	×	UN Children's Fund (UNICEF) and former World Bank staff in Pakistan

* National WASH specialists with no involvement in the South Asia WASH Results Programme (SAWRP)

2.2 Evaluation themes covered with country stakeholders

- **Country context:** other large sanitation and hygiene programmes; government prioritisation; national context.
- **Programme design and functioning:** scale; staff and partners; contracting; theory of change (TOC); implementation activities prior to 2015 and 2016–2017.
- **Monitoring:** outcome targets; indicator definitions; progress monitoring; survey design.
- **Implementation progress and quality of implementation:** progress to date; quality of implementation; variation in quality; district wide focus; handwashing.
- **Inclusion, sustainability, and health:** progress; variations in achievements; challenges; faecal sludge management (FSM).
- **How PbR played out in-country:** verification process; verification indicators; evidence requirements; changes in approach; payment deductions; lead verifier; benefits of PbR; negative consequences.
- **Degree of learning:** approach; innovations and evolutions; lessons.

2.3 Limitations

2.3.1 Data collection

Table 2: Key data collection limitations and implications

Limitation/ issue	Implications and mitigating action
Restricted movement during Pakistan country visit Q1 2018: In November 2017, Plan International was notified by the Government of Pakistan that its permission to operate in the country was not being renewed and that programme operations had to cease within three months. This created considerable uncertainty as to whether the evaluation mission could go ahead. By the time the mission was confirmed it was too late to obtain the no objection certificates (NOCs) needed from government so that the evaluation team could visit IP offices and beneficiary communities. NOCs are required for all international staff conducting work in Pakistan if leaving the major cities (Islamabad, Lahore, and Karachi).	The restrictions meant that the evaluation team were unable to leave the major cities, and were unable to visit IP offices or targeted communities. In agreement with Plan International and WaterAid, the mitigating action was for the key informants to travel to meet the evaluation team in the cities. This enabled the team to meet all key IPs and a number of government counterparts, though it was generally limited to managerial staff; only a few field staff attended.
Outcome data availability at the time of the case study: At the time of the mission, the final verified results for the outcome phase were not yet available and nor were the results of SAWRP's final Sustainability Assessment Framework (SAF) surveys.	The implication for data quality is that at the time of interviewing programme staff the final outcomes were not available and therefore could not be considered in interviewing supplier staff. However, as the midline and endline outcome survey results were very similar, with hindsight this is not a significant issue. That is, the final outcome survey results did not raise issues that would invalidate the findings from interviews or necessitate significant further interviews.
Bangladesh country programme review: The review team did not visit Bangladesh and therefore this part of the evaluation was limited to a review of documents, remote interviews with country programme managers, and interviews with consortium managers and the lead verifier in the UK.	This is a more serious limitation as the views of IP and government staff are not directly captured by the evaluation. This limitation is a feature of the design. In the analysis, careful attention is paid to clarifying where the Bangladesh findings reveal significant similarities or differences between the two country programmes.
Limited view of financial data: Due to the commercial and PbR nature of the contracts, partners were unwilling to share financial information with the evaluation team. Plan explicitly requested the following contractual clause that was subsequently included in the contract[1].	This restriction severely hampers the scope for commenting on value for money beyond that represented by the 'prices' paid by DFID as set out in the contracts. The analysis and discussion of efficiency and value for money aspects is based solely on the qualitative reflections of the supplier staff and discussed in relation to contract value as a whole.
Limited direct engagement with beneficiaries: The evaluation design meant that there was limited third-party data collected.	As part of addressing this limitation the evaluation team considered using some of the beneficiary feedback mechanisms established by suppliers, but no viable options were identified.

[1] The clause reads: "*'the supplier will not be required to report to DFID on expenditures on the project. Where DFID or its MVE contractor wishes to carry out value for money analysis, expenditure data may be requested for this purpose within defined and agreed terms of reference. DFID will not be entitled to carry out an audit of the project based on a schedule of expenditure but may undertake investigations into fraud, bribery and corruption if it so chooses and expect the full co-operation of the supplier."*

2.3.2 Bias

There are several sources of potential bias arising from the data collection. While in all cases mitigating action was taken in the analysis, where possible, these remain important to consider in relation to the analysis:

- **Programme staff interviews:** The vast majority of the analysis is based on interviews with programme staff. This is because the evaluation is explicitly for learning purposes rather than accountability, and as such the experiences of implementers was seen as one of the most important facets. The result is that the majority of the analysis rests on a primary data source that has an incentive to cast the programme in a positive light. While during the analysis a focus was placed on triangulating data from interviews to arrive at the summary conclusions the evaluation team do not seek to question the experiences of the implementing staff as reported, and a focus was placed on accurately reflecting the reported experiences.
- **Limited primary data on results:** The evaluation team did not collect primary data on outputs and outcomes given DFID's investment in the results verification. As such, the verified results data are assumed to be accurate as the evaluation team have limited means to validate these. Thus, should there be any inaccuracies in the results data the analysis based on these data will have these errors embedded within it.
- **Monitoring, verification, and evaluation (MVE) contract:** The verification team and the evaluation component were commissioned under a single contract. However, the verification workstream (led by Itad) and the evaluation component (led by Oxford Policy Management (OPM)) are managed separately; OPM and Itad were contracted jointly as the e-Pact consortium. This is a potential source of bias as regards the evaluation team's judgements relating to the verification framework. Several steps were taken to minimise the risk of this affecting the analysis, most significant of which was the fact that the verification and the evaluation were independently managed workstreams.

2.3.3 External validity/generalisability

The analysis is deeply rooted in the context of the particular PbR modality used. Salient features are: that there was no grant component and payments were only made on the basis of verified results packages/deliverables; there were no upside incentives – only penalties for underperformance; the programme was a DFID centrally managed programme; the programme used a non-governmental organisation (NGO) delivery channel; and the programme had a very tight hard deadline for results to be delivered.

As such, the analysis is best viewed as pertaining not to all forms of PbR contracting but rather to this particular formulation. Throughout this report attention is placed on documenting the contextual factors that affected implementation and how the suppliers operationalised the modality. While there is learning related to the use of PbR contracting for WASH programmes more broadly, the findings are firmly situated in the context of this particular application of PbR. Furthermore, it is worth noting that the three supplier consortia (SAWRP, Sustainable WASH in Fragile Contexts (SWIFT), and Sustainable Sanitation and Hygiene for All (SSH4A)) all had different results packages and verification frameworks arising from how the tenders were formulated and contracts negotiated.

3 The SAWRP consortium

This section seeks to provide an overview of the SAWRP consortium. This section is structured as follows:

- Section 3.1 provides a broad overview of the consortium structure, implementation areas, and the implementation approaches.
  - These facets are explored in more detail as they pertain to country-specific implementation in sections 3.4 and 3.5.
- Section 3.2 provides details of the consortium's contractual targets, how these were translated into payment milestones, and the verification approach and indicators.
- Section 3.3 presents the consortium's achievements.

3.1 Overview of SAWRP

3.1.1 Consortium make-up

The SAWRP I consortium operated in Pakistan and Bangladesh and was implemented by a consortium led by Plan International with WaterAid, Unilever, and Water and Sanitation for the Urban Poor (WSUP) as consortium partners. The Water, Engineering and Development Centre at Loughborough University (WEDC) and Ipsos MORI were advisory/learning partners in the consortium. The programme ran from March 2014 to March 2018, though for Bangladesh only, DFID is funding an extension up to March 2021 for the delivery of additional outputs and outcomes. The extension, known as SAWRP II, falls outside the scope of the evaluation and for the purpose of this report the programme is regarded as completed, meaning activities are described in the past tense.

SAWRP was contracted under Lot B, and therefore the focus of the programme was on rural sanitation and hygiene, although it included a modest water supply component so that small water supplies could be provided in cases of acute water stress that potentially impacted on hygiene. Plan International and WaterAid implemented the rural sanitation, hygiene promotion, and water supply component of the programme. Sanitation promotion was largely based on Community-Led Total Sanitation (CLTS), though there was also a modest sanitation marketing component to encourage and enable low-income households to install (or upgrade to) durable and hygienic improved latrines.

Much of the work in each country entailed the application of established approaches developed under earlier programmes. In Pakistan, Plan International and WaterAid had previously worked together under programmes funded by other donors.

In Pakistan, both suppliers worked via IPs, all of them NGOs, while in Bangladesh only WaterAid worked though IPs while Plan International deployed its own field staff to work directly with local government.

Table 3 lists the SAWRP consortium members, IPs, and targeted locations in each country. In addition, as noted above, cross-cutting support was provided by the advisory/learning partners Ipsos MORI and WEDC.

Table 3: Overview of IPs' areas of operation

Country	Supplier	IPs	Location: District (Province/Division)	Urban/ Rural
Pakistan	Plan International	Lodhran Pilot Project (LPP)	Lodhran (Panjab)	Rural
		National Rural Support Programme (NRSP)	Bahawalpur, Muzaffargarh, Rahimyar Khan (Panjab)	
			Ghotki (Sindh)	
			Mardan (Khyber Pakhtunkhwa (KPK))	
		Association for Water, Applied Education, and Renewable Energy (AWARE)	Umerkot (Sindh)	
	WaterAid	Muslim Aid Pakistan	Rajanpur (Panjab)	
		NRSP	Thatta, Badin (Sindh)	
	Unilever	Idara-e-Taleem-o-Aagahi (IDA)	21 Districts across Pakistan	Mixed
Bangladesh	Plan International	Local government	1 District (Barisal)	Rural Mostly rural, few peri-urban Rural
			2 Districts (Dhaka)	
			3 Districts (Ranqpur)	
	WaterAid	SKC, VERC, ESDO	1 District (Rajshahi) 3 Districts (Rangpur)	
	Unilever	4 activation agencies (Asiatic, Interspeed Market Access, Searchlite)	11 Districts (Dhaka)	
			1 District (Barisal)	

Box 1: Role of the advisory partners

WEDC and Ipsos MORI

The role of WEDC was to coordinate and support monitoring, research and learning with input from Ipsos MORI on data collection formats and data management and analysis. WEDC assisted in:

a) Designing the output/outcome monitoring frameworks and questionnaires. It has tended to focus more on the more challenging outcome surveys;

b) Designing the SAFs and terms of reference (for all four components), providing advice on implementation, reviewing outputs, and developing synthesis reports;

c) Developing the project learning strategy and specified outputs, including supplementary research;

d) General strategic and technical advice to project partners, participating in negotiations and facilitating sessions in workshops, meetings, and Skype calls.

Ipsos MORI was responsible for conducting the baseline and endline surveys for some of the key results (hygiene).

Governance of the SAWRP consortium was provided by an executive team which sat within Plan International UK in London. The team provided overall technical and managerial guidance for the consortium. This team was also supported by WEDC and Ipsos MORI. Plan International UK also employed a dedicated WASH programme manager to provide day-to-day technical and management support to programme directors based in Bangladesh and Pakistan. In both countries, the implementation of SAWRP was supported by a Country Coordination Unit (CCU). In Pakistan this consisted of the programme director, MVE specialist, and finance coordinator. The roles and responsibilities of the CCU were limited to coordination amongst consortium partners. In Bangladesh, this consisted of Programme Director/Head of CCU (1), Head of M&E (1), M&E Manager/Specialist (1) and Finance Manager (1). A key responsibility of the CCU was aggregating the information needed for reporting to DFID on results.

Plan International UK was the contract holder with DFID, and between the three lead partners there was a PbR arrangement – discussed more in sections 4.1.5 and 4.3.1. With the exception of WaterAid Bangladesh's partners, all IPs were contracted using 'normal' grant agreements.

In 2017, Plan's contract was extended until 2021 with an additional £13.5 million in budget allocated to continue the programme in Bangladesh[2]. This extension entails the delivery of new output and outcome results and is commonly referred to as 'SAWRP II'. The scope of the evaluation pertains only to the original SAWRP contract, though some reference is made to changes between SAWRP I and SAWRP II where learning from the initial contract was applied.

3.1.2 Overview of implementation approaches

The WASH Results Programme was divided into two phases: the output phase lasted between 2014 and March 2016[3] and the outcome phase was between January 2016 and March 2018. During the output phase, payments were based on the delivery of output-level results only and similarly during the outcome phase were based only on outcome-level results. In the context of SAWRP the delineation of payments between phases, combined with the tight timelines and hard deadline for output-level results, significantly affected the programme strategy to the extent to

[2] The extension was not taken up in Pakistan – this was the supplier's choice rather than enforced by DFID.
[3] With a one-quarter extension later added – the original timeframe was for the output phase to end in December 2015.

which it is reasonable to characterise the two phases as having separate strategic focuses. It should be emphasised that many of the programme activities were not new to the WASH sector or organisations – rather it was their timing and the emphasis in programming that changed between phases.

Output-phase strategy

In Pakistan, the implementation approach for rural sanitation was predominantly CLTS, though WaterAid Pakistan and Plan International Pakistan had a slightly different focus in programming. WaterAid placed a stronger emphasis on the initial construction of improved latrines (as per UNICEF/World Health Organization (WHO) Joint Monitoring Programme for Water and Sanitation (JMP) criteria) while Plan International focused on the construction of 'basic'[4] latrines in the first instance, with a view to upgrading later in the programme[5].

The implementation approach was similar in many respects in Bangladesh,[6] i.e. predominantly CLTS with some modifications. Plan International Bangladesh placed a stronger emphasis on the initial construction of improved latrines (as per JMP criteria). WaterAid encouraged the construction of 'basic' latrines in the first instance only (1.6% of total) until December'2014 and later moved to improved latrine. Notable features are that the programme included an explicit subsidy component for the very poor[7] – who were identified using Participatory Rural Appraisal type methods – in the initial engagement with communities. A further notable difference is that Plan International Bangladesh implemented the programme in much closer partnership with local government institutions; for all other SAWRP partners, implementation was through NGO IPs in coordination with government.

Unilever's implementation in both countries was largely independent of the two other main consortium partners. The school hygiene promotion component focused on handwashing using their 'School of Five' methodology, which comprised a 21-day intervention package per school. In each country, implementation was led by a single IP. In Bangladesh, WSUP (a British INGO) provided technical support to this component as well as continued to promote hand washing promotion in these schools following 'close-touch' and 'light-touch' approaches during the outcome phase till December 2016. Unilever's implementation ended early in the outcome phase with their final results assessed in Q4 2016.

Outcome-phase strategy

In both countries, WaterAid and Plan International did not clarify their outcome-phase strategy until Q1/Q2 2016. The strategy once developed was most clearly manifested through revised workplans and grant agreements with partners.

In both Pakistan and Bangladesh the overall focus of the programme shifted towards building the capacity of local-level institutions (both government and community-based) with continued follow-up promotion and monitoring. Some degree of output delivery (outside of the payment milestones) also continued in both countries alongside the adoption of new programme elements (particularly

[4] Which may, or may not, have been 'JMP improved'.
[5] For example through the use of more durable materials.
[6] It should be noted that the partners in Bangladesh were only interviewed at endline, and as such the evaluation team have a comparatively less nuanced understanding of output-phase implementation.
[7] WaterAid Pakistan also provided 'demonstration latrines', usually to one or two households also identified during the initial triggering process.

promotion in schools). The outcome-phase strategies in each country are detailed further in sections 3.4.4 and 3.5.4.

3.1.3 Significant changes to design during implementation

The team did not note any significant deviations in approach (e.g. SAWRP locations or targets). However, the following should be noted: i) the definition of the verification indicators and the verification requirements was an ongoing process throughout the course of the programme; and ii) as outlined above, the outcome-phase activities were not well specified until well into the outcome phase.

3.2 SAWRP targets and the verification approach

3.2.1 Targets set in the contract

Table 4 outlines the deliverables as per the contract. Unlike the other two supplier consortia (i.e. SWIFT and SSH4A), SAWRP's deliverables were exclusively related to output and outcome levels (i.e. there were no input or process-related payments). The programme overall set out to achieve just over 1.6 million domestic sanitation and hygiene beneficiaries, with roughly 14% of that number benefitting from access to safe and reliable drinking water. By far the biggest target, however, was for school hygiene promotion: almost 5 million children were to be reached using Unilever's 'School of Five' approach.

Table 4: SAWRP deliverables as per contract annex

Deliverable		Consortium target
Hygiene	Deliverable 3.1 (output)	6,609,569 poor men, women, and schoolchildren across 64 districts reached by handwashing promotion in villages and schools
	Deliverable 3.2 (outcome)*	5,948,612 [90%[1]] poor men, women, and children across 64 districts continue to practice handwashing with soap at critical times
Sanitation	Deliverable 2.1 (output)	1,644,175 poor people across 22 districts have access to new household latrines
	Deliverable 2.2 (outcome)	1,150,923 [70%] poor people across 22 districts continue to use basic or improved latrines
Water	Deliverable 1.1 (output)	227,750 poor people across 22 districts have access to safe drinking water sources
	Deliverable 1.2 (outcome)	204,975 [90%] poor people across 22 districts continue to use reliable, safe drinking water sources
Governance related[2]	Deliverable 4.1	Local WASH governance structures operational in all project locations
	Deliverable 4.2	Regional and local governments serving 22 districts continue to plan, coordinate and monitor WASH activities effectively

Source: Contract annex dated September 2014

[1] Of those reached – NB this was the original target in the contract and was later revised through the definition of the indicators. [2] NB governance-related indicators were included in the annex but were not linked to payment.

Table 5 compares the output-phase targets by country and area. Pakistan targeted a higher number of beneficiaries for sanitation but fewer for community water supply, and substantially fewer for hygiene. This is largely associated with the scale of the school hygiene component in Bangladesh (which was led by Unilever); the target for Bangladesh was more than five times that for Pakistan.

Table 5: Output-phase targets by county

Deliverable		Consortium target	Pakistan	Bangladesh
Output level	Water	227,750	90,000	137,750
	Sanitation	1,644,175	1,000,000	644,175
	Hygiene (Total)	6,609,570	1,786,000	4,823,570
	Hygiene (Household)	*1,644,175*	*1,000,000*	*644,175*
	Hygiene (Schools)	*4,965,395*	*786,000*	*4,179,395*

Source: Contract annex dated September 2014

In the contract annex the outcome-level targets were expressed in beneficiary numbers. As part of establishing the Monitoring and Verification Framework for Outcomes (MVOC) the hygiene targets for outcomes were revised to percentage point (pp) increases over baseline across three components. The indicators and targets were different for school and household level hygiene promotion. Table 6 outlines the outcome-level targets for household and school hygiene promotion.

Table 6: Targets for hygiene outcomes

Promotion level	Component indicator/target	Component	Pakistan		Bangladesh	
			Baseline result	Target increase	Baseline result	Target increase
Household	'At what points in the day do you wash your hands with soap?' Percentage of respondents saying 'before eating' increases by x percentage points	Reported practice	74%	10pp	45%	15pp
	The interviewers should observe if soap or a soap substitute is available at their handwashing facility. At least 10 pp more of these interviewer observations should reveal the presence of soap or a soap substitute, compared to baseline.	Observed behaviour	67%	10pp	51%	10pp
	Percentage of people who can name three or more critical times for handwashing increases by 15 pp at endline compared to the baseline.	Knowledge	36%	15pp	17%	15pp
School	Percentage of respondents able to state all the five times increases by 10pp	Knowledge	54%	10pp	46%	10pp
	Percentage of respondents saying 'every day' increases by 10pp	Reported practice	55%	10pp	45%	10pp
	For children practising fewer than six steps of quality handwashing at baseline (i.e. five or fewer), the mean number of quality handwashing steps they undertake at mid/endline will increase by +1 step	Observed behaviour	3.6 steps	+1 step	3.8 steps	+1 step

Source: Monitoring, Verification and Evaluation Framework for Project outcomes (MVOC) SAWRP, 2016 (updated November 2016)

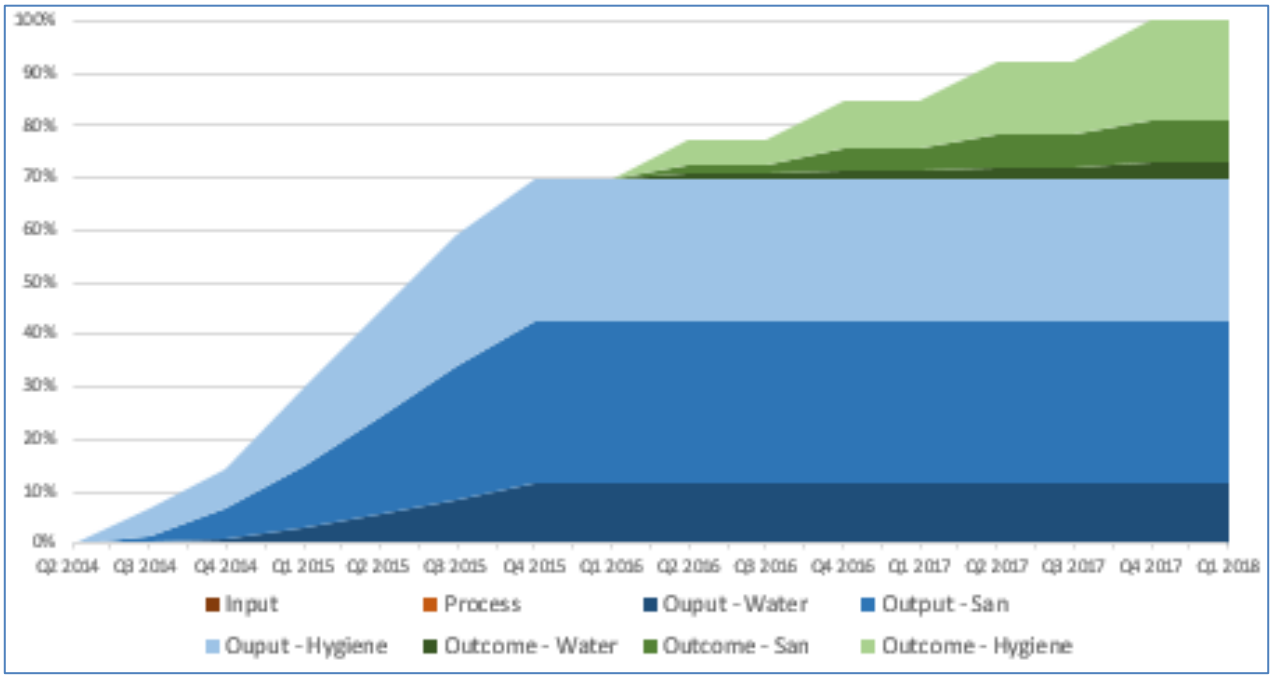
3.2.2 Results linked to payment (payment milestones)

SAWRP was contracted under Lot B, and therefore the focus of the programme was predominantly on rural sanitation and hygiene. This is reflected in the payment structure under the contract (see Figure 1). It is worth noting that SAWRP payments were exclusively related to programme output and outcomes[8]. One consequence is that the verification framework was less complex than the

[8] This was unique to SAWRP. The SSH4A and SWIFT milestone payments included some related to inputs and processes

other two suppliers' frameworks in terms of the number of indicators to be verified – something discussed later in this section.

Figure 1: Structure of payments over the course of the WASH Results Programme for SAWRP



Sources: Analysis is based on the contract annex dated September 2014. The coding of the data by results area is by the authors.

Figure 1 presents an analysis of the contract structure with payment milestones classified[9] by inputs, processes, output-, and outcome-related payments. Over the four years of the programme, approximately 70% of the payments were planned to be made in the output phase (April 2014–December 2015[10]), and the remaining 30% in the outcome phase.

SAWRP was paid at 10 payment points, with six of the verification/payment points in the output phase and four in the outcome phase. In practice, the number of payment points was fewer than was originally envisaged in the contract (15 payment points). All deliverables were independently verified by the e-Pact MV team. During the outcome phase the nationwide government census in Pakistan meant that one planned payment round could not take place; this is discussed further below. Beyond the reforecasting there were no significant changes to the beneficiary numbers or locations across the course of the contract.

[9] By the authors.
[10] Later extended by one quarter.

Table 7: SAWRP payment/verification rounds

	2014			2015				2016				2017			
	Output-related payments							*Outcome-related payments*							
Calendar quarter->	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Verification of SAWRP deliverables		X	X	X	X	X	X				X	X	X		X

Source: MV verification reports

Sources: Analysis is based on the contract annex for the supplier. The coding of the data by results area is by the authors.

Eligibility for payment at each milestone was independently assessed by the e-Pact MV team. Deliverables linked to payment were all expressed in beneficiary numbers (see Table 4 above). In addition, the SAWRP contract included some governance-related indicators that were not linked to payment. The evaluation team are unaware of reporting against these contract milestones via the verification process.

The definition of the payment indicators was established through the process of agreeing the verification frameworks between SAWRP, the MV team, and DFID. In the case of SAWRP, during the output phase 'Monitoring and Verification Frameworks for Project Outputs' (MVOs) for Pakistan and Bangladesh were developed. The process for developing these is discussed further in Section 3.2.4. The MVOs outlined the steps the partners were required to take in data collection, filing, internal quality assurance, data consolidation, and analysis to ensure that the results reported were accurate and transparent. The MVOs also specified how these would be verified. During the outcome phase, the MV team in partnership with SAWRP and DFID developed the MVOC. As with the MVOs in the output phase, the MVOC specified the requirements for data collection, the verification process, and the outcome-level indicators related to payment. Unlike the MVOs there was a single MVOC for all of SAWRP. In both cases, the MVOs and the MVOC translated the broad consortium targets into more specific and measurable targets, in effect refining the targets. A near final version of the MVOC was presented to DFID in Q1 2016, and the first round of outcome surveys was conducted shortly after that, although the MVOC was revised in late 2016.

3.2.3 Verification process

As with all the suppliers the verification process was built around the supplier's existing monitoring frameworks, albeit with the MV team requiring additional internal quality assurance processes and evidence as part of the results packages. Box 2 outlines the common elements of the verification framework under the WASH Results Programme.

Box 2: Common elements of the verification framework across suppliers

Due to the PbR financing modality of the programme, suppliers were only paid for results that have been independently verified. Therefore, the verification process, which confirms whether or not the supplier has delivered the agreed results, is a central element of the programme. The independent verification of suppliers' results is based on a **systems-based approach**. This approach was not specified in the terms of reference but was chosen by the MV provider to match the budgetary envelope of the terms of reference.

Systems-based verification means that evidence regarding the achievement of results is not established through independent data collection by the verification agent but is instead based on data generated by the supplier's internal monitoring and reporting systems. This implies that a strong focus of the verification process is appraising the robustness of the supplier's internal monitoring and reporting systems. The MV provider set out their approach for verification at the end of the Inception Phase in September 2014. The approach is based on three core elements:

- **A systems appraisal of the supplier's internal monitoring and reporting systems:** A comprehensive systems appraisal is conducted ahead of the first full verification cycle to map which internal monitoring and reporting systems will generate the evidence needed for verification. If systems are deemed to be insufficient, corrective action is recommended by the verifiers. The systems appraisal is repeated ahead of each verification cycle until the systems are deemed to meet the required standard for evidence generation, and/or if evidence requirements change over time.
- **Desk-based verification of supplier-generated evidence:** First, a list of evidence requirements is drawn up by the MV team (the 'Form 2'), tailored to each verification round as necessary. This evidence is then submitted by each supplier and checked for completeness by the MV team.
- **Field-based verification using MV team-generated data:** in parallel to the desk-based verification of evidence described above, the MV team carrying out field visits double-checks the veracity of evidence submitted and the quality of results achieved, and assesses the likely sustainability of results achieved.

These three elements take place in parallel and inform the **conclusion** by the verification team regarding whether a given supplier has delivered the agreed results (the quarterly **verification report**). This conclusion is passed on to DFID, who make a **payment decision** on the results to be paid for that quarter. An **After-Action Review** is frequently held thereafter to identify lessons and agree on actions to take in forthcoming verification rounds.

All verification forms (Form 2s) are built around the same elements: for each verification indicator, the form specifies the indicator definition, data source and data requirements for suppliers, and the methodology of analysis for the verifier. Indicators usually include a numeric assessment of the number of results achieved and a list of the documentation required to establish the veracity of the result and also its quality, if applicable. Each Form 2 also includes the methodology for how a payment decision is made, based on the aggregate analysis of all the indicators pertaining to that deliverable, such as whether payment is proportional to the number of results achieved or whether it is made based on a pass or fail.

To ensure that the verification process was feasible and appropriate, the evidence requirements set out in the Form 2s were tailored for each supplier, and for each results deliverable, and in some cases even tailored for different countries or IPs. This resulted in 42 different Form 2s[11] for the output phase. As the verification methodology was designed and adjusted while supplier implementation activities were ongoing, several modifications were made to the evidence requirements during the output phase.

During the output phase, results were assessed separately for each of the consortium members against their particular set of payment triggers. However, in the outcome phase, when verification was based on surveys looking at programme results overall, payments were made to the

[11] Across the three suppliers.

consortium lead then divided up by the consortium in proportion to each member's contribution to the results.

The outcome-phase surveys were implemented in a random sample of project communities. For the water and sanitation outcomes the partners implemented the surveys (using the programme field staff as enumerators), while the hygiene results were assessed via a survey implemented by Ipsos MORI. Table 8 outlines the envisaged timing of these surveys at the outset of the outcome phase. It should be noted that the March 2017 water and sanitation outcome surveys could not be implemented due to the Government of Pakistan conducting the census and calling a moratorium on all other survey data collection in the country during that period.

Table 8: Timing of the outcome-phase surveys

Result	Survey name	Survey implementer	Survey delivery (planned)	Report and invoice submission (planned)
1.2 Water	Outcome survey – water	SAWRP partners	Aug-16	Feb-17
			Mar 2017	Sep-17
			Dec-17	Mar-18
2.2 Sanitation	Outcome survey –sanitation		Aug-16	Feb-17
			Mar 2017 (P) Jun-17 (B)	Sep-17
			Dec-17	Mar-18
2.2.1 Hygiene, household – (WaterAid and Plan)	'Household survey'	Ipsos MORI	Oct 2016	Jan-17
			Sep / Oct 2017	Jan-18
2.2.2 Hygiene, school – (Unilever)	'Child Survey'		Dec-15 (P) Apr-15 (B)	Oct-16
			Dec 2015 and Mar 2016 (P) Sep-15 (B)	Oct-16

Source: Adapted from the MVOC

3.2.4 Verification and payment indicators

For SAWRP, each verification round consisted of a repeat assessment of the relevant output or outcome indicators. At the output and outcome levels, Unilever was assessed by different indicators (and therefore Form 2s) to WaterAid and Plan International. For Plan International and WaterAid, at the output level there were separate Form 2s for each country and at the outcome level there was a single Form 2 covering both partners.

Output-phase indicators

One simplifying aspect of the verification process in Pakistan was that DFID accepted SAWRP's proposal that each sanitation beneficiary would also be counted as a hygiene beneficiary[12], provided the MV team was satisfied that all planned hygiene promotion interventions had taken place in each targeted community.

Outcome-phase indicators

In the outcome phase, Plan International, WaterAid, and Unilever received payments based on three targets (detailed in Table 9) for hygiene, with the targets varying slightly by IP. Further details are provided below.

Table 9: Outcome-phase targets

Area	Indicator	Measurement
Water	90% of poor people across the project districts continue to use reliable, safe drinking water sources	Red/Amber/Green (RAG) ratings based on reported water availability throughout the year and observed
Sanitation	75% of poor people across project districts continue to use basic or improved latrines	RAG ratings based on self-reported frequency of toilet use and observed toilet condition
Hygiene	The pp increase of poor men, women and children continue to practice handwashing with soap at critical times	Increase in knowledge of all five critical occasions
		Self-reported practice
		Knowledge of critical times (for Plan and WaterAid). Observation of behaviour (for Unilever)

The outcome-level indicators for SAWRP were constructed by aggregating RAG ratings based on the survey data. In all cases the indicator was passed if the aggregated RAG rating was Amber or Green.

- **For water:** for each sampled water point, three beneficiary households were selected from the output database for survey.[13] In addition to interviews, the enumerator assessed whether the water point was an 'improved source' and whether water could be drawn at the time of the visit. Both elements of the RAG rating (i.e. self-reported use and observed use) must be classified as Green (water available for 12 months of the year) for the combined response to result in a payment trigger.
- **For sanitation:** For each sampled latrine the household was interviewed. The enumerator first confirms that it is an improved latrine. If it is unimproved, such as a pit latrine without slab, and open pit or hanging latrine, or open defecation, this would automatically result in a Red rating. If it is an improved latrine, then a two-component RAG-rated indicator comprising self-reported use and observed was used. A Red rating for 'observed use' automatically results in an overall Red rating. Self-reported use 'rarely or not at all' will result in an overall Red rating, whatever the observed use. Amber for reported use was if the household responded that the latrine was

[12] It should be noted that the verification of hygiene promotion activities formed part of the verification framework – though the verification reports consistently raised issues surrounding the quality of the evidence submitted. The verification reports of Q4 2014, Q1 2015, Q2 2015, Q3 2015, and Q4 2015 all raised varying issues regarding the verification of hygiene promotion activities across both countries

[13] Five households are randomly selected per water point to allow for refusals/non-availability.

used 'Most of the time' or 'don't know'. Both elements of the RAG rating (i.e. self-reported use and observed use) must be classified as Green or Amber for the combined response to result in a payment trigger.

- **For hygiene:** At the end of the output phase the numbers of beneficiaries who have benefited from hygiene interventions was fixed. For the outcome phase, hygiene improvements were assessed using a random sample drawn from areas of project intervention. Progress towards the hygiene target was assessed using a composite indicator made up of three components which focus on handwashing behaviour, reported practice, rapid observation and knowledge. The indicators used were presented in Table 6 in Section 3.2.1.

3.3 Achievements

3.3.1 Table of aggregated results by country

Table 10 provides an overview of the output-phase results for SAWRP in Pakistan and Bangladesh. On all indicators in both countries there was substantial overachievement against the target. Overachievement was not paid for, but is included here to illustrate performance.

Table 10: Output-phase results

Deliverables		Bangladesh			Pakistan		
		Target	Actual	% of target	Target	Actual	% of target
1.1	The number of poor people having access to safe and reliable drinking water sources	137,750	169,706	123%	90,000	110,341	123%
2.1	The number of poor people having access to basic or improved household latrines	644,175	735,405	114%	1,000,000	1,205,582	121%
3.1	The number of poor men, women and schoolchildren that have been reached by handwashing promotion in villages and schools	4,823,570	5,100,766	106%	1,786,000	2,020,793	113%

Source: MVOC, December 2016 version.

3.3.2 Outcome-phase results

Table 11 presents the verified results of the Q1 2017 and Q1 2018 outcome assessments from the 'Outcome surveys' and the 'Household Survey' (hygiene). These results indicate substantial overachievement against the targets by Q4 2017. In the case of water, 100% of beneficiaries were continuing to use improved water source[14] at the time of the survey based on the RAG rating and use reports. In the case of sanitation, 99% of the sampled toilets were classified as improved,

[14] This could be a SAWRP source or other improved source

showed evidence of use, and households reported use. The picture was more mixed for hygiene, though there was significant overachievement on the self-reported practice and observation of a facility.

Table 11: Verified outcome-phase results of SAWRP

Area	Indicator and target	Pakistan		Bangladesh		
		Q1 2017	Q4 2017	Q4 2016	Q2 2017	Q4 2017
Water	90% of poor people across the project districts continue to use reliable, safe drinking water sources	100%	100%	97.1%	99.2%	98.6%
Sanitation	75% of poor people across project districts continue to use basic or improved latrines	96.1%	99%	99%	99.1%	99.2%
Hygiene	Reported practice: Percentage of respondents saying 'before eating' increases by: 15pp(B)/10pp(P)	+9pp (81%)	+33pp	+32pp (77%)	*n.a.*	+38pp
	Observation: At least 10pp (B&P) more of these interviewer observations should reveal the presence of soap or a soap substitute, compared with baseline	+7pp (74%)	+27pp (74[15]%)	+27pp (78%)	*n.a.*	+32pp
	Knowledge: The percentage of people who can name three or more critical times for handwashing increases by 15pp (B&P) compared to the baseline	+2pp (37%)	+11pp	+9pp (26%)	*n.a.*	+20pp

Source: Initial submissions of Q1 2017, Q2 2017, and Q42017 verification reports

[15] N.b. the progress levels are different despite the same survey results. This is as during the Q1 2017 survey beneficiaries and non-beneficiaries were interviewed. This is different from the Q4 2017 survey where only beneficiaries were interviewed.

Table 12: Child survey results (Unilever)

Area	Indicator and target	Pakistan*		Bangladesh	
		Midline	Endline	Midline	Endline
Hygiene (Child)	Knowledge: Percentage of respondents able to state all the five times increases by 10pp	-39% (15%)	-27pp (23%)	+16pp (62%)	+33pp (79%)
	Reported practice: Percentage of respondents saying 'every day' increases by 10pp	+19pp (74%)	+20pp (80%)	+28pp (73%)	+43pp (88%)
	Observation: For children practising fewer than six steps of quality handwashing at Baseline (i.e. or less), the mean number of quality handwashing steps they undertake at mid/endline will increase by +1 step	+2.3 steps (5.7)	+3.0 steps (6.6)	+3.1 steps (6.9)	+4.0 steps (7.8)

Source: Q4 2016 verification report

*Note: The baselines for the mid and endline survey analysis in Pakistan are different because they reflect attrition rates during the surveys. The child surveys are longitudinal cohort studies and the sample size during the mid and endline was reduced if the same children could not be found.

3.4 SAWRP in Pakistan

Table 13: Summary of respondents in Pakistan

Stakeholder	Details	Comment
Country verifier	Interviewed	The interviewee was appointed in late 2017 and therefore had limited programme knowledge
Consortium partner staff	Plan International and WaterAid only	The team could not travel outside of Islamabad and Karachi due to security restrictions. Most of the supplier staff met were managers, as field staff were no longer in post. Unilever's involvement in the programme ended long before the visit and relevant staff were not available for interview
IPs	Met all IPs: NRSP, AWARE, MAP and LPP	Again, the team mostly met project managers, as field staff were no longer in post[16]
Government counterparts	Interviewed a number of Public Health Engineering Department (PHED) and Local Government Department (LGD) managers from KPK, Panjab, and Sindh	All senior provincial respondents. Interviews with federal ministry officials were not prioritised as WASH is a provincial subject in Pakistan. However, other sector experts provided some context on sector status at national level
Other sector experts	UNICEF and former World Bank WSP staff	Two UNICEF WASH specialists were met, working at national and provincial level

3.4.1 Country context

Progress in WASH

When programme operations started in 2014, the UNICEF/WHO JMP update showed that Pakistan was significantly off-track against the sanitation Millennium Development Goal (MDG) target of 64%, but on track for the water supply target of 93%. Tables 12 and 13 show that, by 2015, the sanitation target had just been met and the water supply target almost met, although total access to improved water supply had not increased over the previous five years. In rural areas, access to improved sanitation had reached 51% and, while this was still low, it represented a 10% increase over the 2010–2015 period.

Although not reflected in the JMP data there are major disparities between locations, with the lowest access rates in the poorest and most remote districts. It is also common in many parts of the country for extended families to live in one compound and share a single toilet. Against this backdrop, meeting the Sustainable Development Goals (SDGs) in Pakistan will be a huge challenge.

[16] Some former field staff (social organisers) were consulted for the sanitation RCT survey and the findings are an additional point of reference where available.

Institutional context

There is currently no national water supply or sanitation policy or strategy, and no sector monitoring system. Instead, each province is responsible for making its own arrangements within a decentralised government framework. This said, there is a *de facto* rural sanitation strategy known as Pakistan Approaches to Total Sanitation (PATS). Based on CLTS but with some additional components, PATS emerged from a programme known as Rural Sanitation in Flood-Affected Districts (RuSFAD), which began in 2010 following major floods that year. Both Plan International and WaterAid played a role in the programme, which was led by UNICEF.

PATS is now followed by most players in the sector (within and beyond government), although there are some minor differences in approach. Both organisations include an element of sanitation marketing in their approach, whereby selected small businesses offer affordable latrine components and skilled labour for the construction of improved facilities.

Pakistan has a federal government system and WASH is a provincial subject. However, lead responsibility for WASH at sub-national level is not clearly defined. PHEDs have for decades played a lead role in the development of new rural water supply schemes (although they have been less active in supporting operations and maintenance) but until recently were not active players in sanitation. Meanwhile, there is broad consensus that, following decentralisation, local governments should take on responsibility for WASH, or at least for supporting the long-term use and functionality of facilities post-installation, but there is a long way to go. The anticipated role includes a role in post-Open Defecation Free (ODF) follow-up to encourage long-term latrine use, including the maintenance, repair, and upgrading of existing toilets. Local health departments, via lady health workers, already have a role in promoting sanitation and hygiene, but the extent to which this actually happens varies enormously, being more common in the context of externally supported projects. UNICEF, while working closely with the PHEDs, is helping to develop the role of local governments at district and union (sub-district) level, particularly in WASH planning and coordination and providing a supportive environment for sustainability.

Government implementation capacity in WASH has been weak for many years – particularly in sanitation and hygiene promotion – and external agencies have relied heavily on NGO partners. A notable exception here is Panjab, the only province where PHED has complemented externally funded projects by developing its own PATS programme and investing some $4 million of government funds in the initiative. In Sindh, a large WASH programme is about to begin, which is funded through a multi-donor trust fund coordinated by the World Bank: the Multi-Sector Action on Nutrition (MSAN) programme will run in Sindh with $26 million in World Bank funding until 2020. Though it is multi-sector in name, the focus is overwhelmingly on sanitation[17].

Socioeconomic and cultural context

In many programme areas, it is common for an extended family – which can be very large – to share a single compound with just one toilet. Sometimes this means that dozens of people share the facility, and this could result in some family members defecating outside to avoid queueing or to stop the toilet filling up too quickly. This is significant in the context of the SAWRP programme as sanitation and hygiene beneficiaries were calculated on the basis of using multipliers for the number of people in a household. Where the numbers of people sharing is large, simply having one toilet per compound may be insufficient provision for the achievement of ODF status. Some

[17] Details are available here: http://projects.worldbank.org/P158769?lang=en [accessed July 2018].

programmes therefore set a maximum number of users per facility for the purposes of counting beneficiaries for payment. In the case of SAWRP this cap was set at 14 people, i.e. no one new latrine could have more than 14 beneficiaries.

Physical and environmental context

In recent years, Pakistan has been subject to repeated humanitarian emergencies, not least the devastating floods of 2010. Many of the locations where development agencies work have been affected by emergencies and occasionally still are. As a result, external support has fluctuated between disaster relief and support with a longer-term perspective; this partly explains why hardware subsidies remain a feature of many sanitation and hygiene projects despite the adoption of CLTS.

Some SAWRP project locations were flood-prone and while none were affected by flooding on the scale seen in 2010 there were some damaging flash floods over the course of the programme, particularly in parts of Panjab. In contrast to this, Sindh is more arid and subject to greater water stress than the other locations where SAWRP operates.

3.4.2 Scope and scale of SAWRP in Pakistan

The core of the SAWRP programme in Pakistan is CLTS and hygiene promotion in rural areas led by Plan International and WaterAid and implemented via four partner organisations: NRSP (a large national NGO), Muslim Aid (an INGO), plus AWARE and LPP, both local NGOs. Linked to this component, the programme has also provided a limited number of water points (typically shallow boreholes with handpumps or water treatment facilities) as an ODF reward in villages with acute water supply problems.

Within Pakistan, SAWRP operated in a total of 12 districts: Plan International was active in nine districts, while WaterAid works in three districts. Plan International Pakistan created three regional offices – in Bahawalpur (Panjab), Mardan (KPK), and Umerkot (Sindh) – to support implementation during the output phase. Each office employed a project manager, a CLTS trainer, a technical specialist, and a monitoring, evaluation, and reporting specialist. Links were also created to other WASH programmes, such as the DFID-funded Accelerating Sanitation and Water for All programme (ASWA) managed by UNICEF. The provincial offices were dissolved at the end of the output phase and staff redeployed into government structures at district and provincial level. WaterAid had provincial representatives in both Panjab and Sindh, and while these were not appointed to support SAWRP specifically they were available to undertake SAWRP-related tasks when the need arose.

At the planning stage, the Pakistan programme targeted districts where a high proportion of residents were poor and baseline levels of latrine coverage were very low (or assumed to be very low), and which had not been served by other WASH programmes in recent years. One of the early challenges faced by IPs was finding enough beneficiaries in their assigned locations due to higher-than-anticipated baseline coverage: a baseline survey found that 62% of households in programme areas were using either private latrines or sharing with one other household[18] and that 60% of these latrines were improved. To meet programme targets, which were defined in terms of beneficiary numbers, SAWRP expanded its geographical coverage, for which larger field teams

[18] South Asia WASH Results Programme: Baseline Survey Report – Ipsos MORI (February 2015).

and additional transport were required. As a consequence, IPs were obliged to renegotiate certain financial aspects of their agreements with consortium partners.

3.4.3 Suppliers and IPs

Plan International Pakistan and WaterAid Pakistan have an established working relationship built over several years. PIan International was lead IP in the RuSFAD programme from 2010 to 2012, while WaterAid was a third-party monitor in the same programme. Box 3 provides more details on the RuSFAD programme and PATS.

Much of the work undertaken by SAWRP in Pakistan represented the extension, expansion, or replication of recent or ongoing projects managed by WaterAid and Plan International. Unilever was already operating at scale in Pakistan when SAWRP began, and the WASH Results Programme has enabled the programme to expand to reach a further 750,000 children. The 'School of 5' approach involves a 21-day intervention in each school, based on a methodology developed in collaboration with the London School of Hygiene and Tropical Medicine.

Box 3: The PATS strategy

The PATS strategy was approved in March 2011. PATS works towards achieving and sustaining ODF environments in rural and urban contexts with clear emphasis towards behaviour change and social mobilisation. The approach endorses the use of a number of total sanitation models, including CLTS, School-Led Total Sanitation, Component Sharing[19] , Sanitation Marketing, and Disaster Response.

PATS emphasis using a 'Total Sanitation Approach' to change behaviours by stopping open defecation and achieving 100% ODF status at community level. The approach then emphasises the need to stimulate and sustain demand at community level to achieve the remaining 'total sanitation' outcomes, i.e. sustainability of an ODF environment, consistent use of sanitation facilities, affordable and informed solutions, promotion of hygiene, provision of drainage, wastewater management, and solid waste management.

Plan International Pakistan and WaterAid Pakistan's IPs were all NGOs, and in all cases the IPs were contracted via grant agreements as opposed to being contracted using a PbR modality. Both assigned full-time staff to support the implementation at the district level. IPs have considerable experience in community-based development approaches, in particular the use of the CLTS approach in rural WASH programmes. IPs also have wider experience in multi-sector programmes such as education, health, livelihoods, and microfinance, creating possible links to other poverty-reduction programme activities. Table 14 provides an overview of the SAWRP partners in Pakistan, their implementation areas, and a brief summary of their work in WASH prior to the WASH Results Programme.

Some of the IPs had been implementing CLTS in the SAWRP districts for a year or more prior to SAWRP and had established working relationships with government agencies at district level, with Union Councils (UCs), and with many of the targeted communities. Not all of the IPs had a track record in CLTS, however, or in WASH generally outside of emergency contexts. For example, Plan International's partner AWARE had limited previous experience, and WaterAid's partner Muslim Aid were implementing CLTS for the first time. In these cases, the country partners provided not

[19] A project based on a component-sharing approach comprises a set of discrete activities (components), allowing the community to independently undertake one or more of these activities.

only orientation and training but also arranged for more experienced partners such as NRSP to play a mentoring role.

Table 14: SAWRP partners and intervention areas in Pakistan

Consortium partner	IP	Programme area
Plan International	LPP	Lodhran District, Punjab
	NRSP	**Punjab:** Bahawalpur, Muzaffargarh, and Rahimyar Khan Districts
		Sindh: Ghotki District
		KPK: Mardan and Swabi Districts
	AWARE	Umerkot District, Sindh
WaterAid	Muslim Aid Pakistan	Rajanpur District, South Punjab
	NRSP	**Sindh:** Thatta and Badin Districts
Unilever	ITA	21 districts

At the field level, IPs work through implementation staff called 'social organisers' in implementing the intervention. A social organiser carries responsibility for implementation in a number of UCs and the villages/communities within those UCs. Each SO supervises and works with a network of community resource persons (CRPs) who implement promotion activities at the community level together with the social organisers. Social organisers are also the key field-level staff responsible for monitoring and reporting. The CRPs are from the villages and are volunteers, while the social organisers also manage the relationships with the CRPs in their areas.

3.4.4 Strategy changes between programme phases

Implementing a programme with distinct output and outcome phases was a departure for the country partners and, following the completion of the output phase, Plan International and WaterAid spent several months mapping out the way forward. Field work did not come to a halt during this time; IPs maintained contact with targeted communities and continued to monitor progress, but no new initiatives were launched until the MVOC was adopted in early 2016.

Once the strategy and operational approach were clear, and the detailed workplans were agreed with IPs, the focus of programme efforts underwent a distinct shift from expanding latrine coverage and (on a much smaller scale) water supply access to consolidating the use and maintenance of WASH facilities and creating an enabling environment for sustainability by strengthening the supportive role of local government, community structures, and (to a lesser extent) the private sector through the supply of affordable sanitation goods and services. In most cases, there was a reduction in the number field staff between the phases.

Prior to SAWRP, Unilever through its brand 'Lifebuoy' adopted an ambitious strategy to reach more than 1 billion poor people across Asia, Africa, and Latin America by 2015 with handwashing promotions. Unilever was already implementing the 'School of Five' approach at scale in Pakistan and Bangladesh when SAWRP began. The Unilever-led component operated at a faster pace than the rest of the programme and on a different timescale, since it ended on completion of the output

phase. Overlap with communities targeted by Plan International and WaterAid was kept to a minimum to avoid the risk of double-counting beneficiaries. In the event, geographical overlap was limited to parts of just four districts. In these cases, the different sub-components were meant to be mutually reinforcing rather than duplicating efforts, given that Unilever only worked in schools while Plan and WaterAid targeted communities.

In the Outcome phase, the focus of programme efforts shifted from the installation of facilities to consolidating behaviour change and securing an enabling environment for sustainability (see Box 4). Respondents from both lead suppliers and their partners said that they had not anticipated such a radical change in focus and that it stimulated them to review, improve, and intensify their approach to behaviour change communication (BCC). Hygiene promotion had not been absent during the output phase but at that stage involved a lot of mass communication to promote collective action in response to CLTS triggering. In the outcome phase, the focus was more on interpersonal communication, with the intention to visit every household at least once per quarter.

As part of a global WaterAid initiative on hygiene promotion, WaterAid Pakistan adopted a completely new BCC strategy which, amongst other things, included a promotional campaign known as *Behtreen Maan* (Ideal Mother), which targeted women as change agents. While the promotion of handwashing with soap was a central objective, the programme also encouraged the safe storage and consumption of water and keeping toilets clean and functional.

Box 4: Community-level outcome-phase activities by supplier

Key Plan International programme activities

- **Follow-up visits at the village level** in its eight districts to ensure sustainable usage of the latrines constructed during the output phase. During these visits, household latrines were visited and health and hygiene messages were also reinforced. There was also a sanitation marketing component in some programme areas.
- **ODF certification** and ODF celebrations at UC level.
- **Conducting health and hygiene sessions** to reinforce the hygiene messages amongst beneficiaries. Complemented by community-wide activities and some mass media, such as street theatre, sports galas, radio programmes, etc.
- **Water point follow-up visits** were conducted to ensure functionality and usage of water points constructed during the implementation phase.
- **World Toilet Day and Global Handwashing Day events** in project districts to inculcate the importance of sanitation and hygiene. The World Toilet Day events focused on the theme of 'Toilets and Jobs'. The Global Handwashing Day theme of 2016 events was 'Make handwashing a habit'.
- **Follow-up sessions were conducted in government schools**.

Key WaterAid programme activities

- **Health and hygiene sessions (community-level meetings)** to extend awareness-raising efforts regarding personal, domestic, and environmental hygiene conditions. WaterAid introduced a new version of health and hygiene awareness sessions, which had a very systematic approach to include the maximum numbers of a community.
- **Village committee meetings** to plan, meet the community, and mitigate the sanitation-related issues of communities.

- **CRP meetings** (monthly) to review their progress as per their routine workplans and assigned tasks.
- **Sanitary entrepreneurs meetings** were organised to ensure the provision of low-cost sanitary goods at community level so demand could be met on time.
- **ODF follow-up visits** were conducted to maintain the ODF status of certified villages and to mitigate slippage on open defecation-related behavioural issues.
- **Global Hand Washing Day and World Toilet Day** were also celebrated within different pockets of communities to highlight their importance.
- **WaterAid also organised street theatres** in the community, providing a unique infotainment kind of methodology that not only amused the audience but also triggered them to fulfil their basic hygiene-related responsibilities.
- **Radio jingles** were aired through different FM channels, along with **newspaper messages** published in local and national-level newspapers.
- **Quarterly broad-based community meetings** at community levels (this also included village walks by the teams and community together).
- **Positive reinforcement through 'Ideal Mother Groups'** at household and community levels.

There was strong consensus among respondents that activity in the outcome phase was different to that in the output phase but no less intensive; in fact, Plan International cancelled an anticipated reduction in partner field staff while WaterAid delayed it by one year. There were changes in staff deployment, however: Plan International closed its three regional project offices and reassigned the staff to support provincial government agencies in Panjab and KPK.

While BCC was a core component of the outcome phase, it was not the only activity. The programme reported that the vast majority of targeted communities became ODF – a remarkable result by global comparisons – and in fact the bulk of ODF certification happened during the outcome phase. Partners said that most targeted communities were already ODF by the end of the output phase, but it took time for government agencies to complete the formal certification process. For Plan International, the outcome phase also involved encouraging and supporting households which had built basic latrines in response to CLTS triggering to upgrade to improved, and more durable, models. Ensuring easy access to affordable materials and skilled labour via sanitary marts (at least one per UC) was part of this initiative. To encourage the establishment and active operation of sanitary marts, Plan provided the operators with finance on very favourable terms. WaterAid also facilitated the establishment of sanitary marts (though without financial assistance) but tried to ensure that households built improved latrines from the start.

Respondents mentioned a number of other interventions to create an enabling environment for sustainability, though no clearly defined strategy emerged. Initiatives included, for example, the following:

- IPs encouraged UC secretaries to play an active role in post-ODF follow-up to consolidate community commitment to using and maintaining toilets. To this end, field staff would often invite the UC secretary to join them when making community visits. This was partially successful in that many secretaries took an active interest in the project while the IPs were

present, although none adopted funded action plans for longer-term community support and monitoring once the outcome phase was over.

- At district level, IPs further encouraged government leadership in WASH by being active members of multi-stakeholder core groups/WASH forums and encouraging the formation and operation of district ODF committees. Some also facilitated the formulation of district ODF plans. Where the IP had an established presence in a district, some of this work was a continuation of initiatives that pre-dated SAWRP.
- Plan International and WaterAid also provided technical support and advocacy for action on sanitation at provincial government level, though again this was a long-term initiative and not related to SAWRP specifically.

3.4.5 Approach to sustainability

While the sustainability of services is by definition dependent on what happens after the programme ends, the approach to sustainability by suppliers was largely reflected in the focus of their outcome-phase activities. The focus of the effort of suppliers in the outcome phase had three main elements: i) continuing household-level promotion activities to try and instil behaviour change; ii) support and guidance to local government stakeholders for continuing the work; and iii) intensive monitoring of functionality with remedial action take when needed.

Apart from outcome surveys linked to verification, Plan and WaterAid also adopted a SAF, under which additional surveys were conducted to track whether enabling conditions for sustainability were in place. These are discussed further in Section 3.4.6 below.

In the case of water points provided under the programme and as ODF rewards, the approach to sustainability was predicated on community management, with little or no support expected from government. Partners indicated that the shallow handpumps installed were of a type that was very common in programme areas and easy to repair, with spare parts readily available in local markets. The water treatment units provided in some locations were slow sand filters. IPs provided maintenance training and technical support (if required) up to the end of the outcome phase.

3.4.6 Supplier monitoring approach and verification

Verification of the outcome indicators was based on the results of two rounds of surveys. The original plan was to do three rounds, but government did not allow the second to go ahead as it coincided with the national census. In the event, the first was conducted in mid-2016 and the second in late 2017. Each round was in two parts:

a. A survey to assess the functionality and use of sanitation and water supply facilities. This was managed by the CPs with IP field staff serving as enumerators

b. A survey by Ipsos MORI on handwashing practices, based on household interviews. The company appointed their own enumerators.

At the time of the evaluation visit, verified results from the second round of outcome surveys were not yet available but CPs indicated that the findings were better than for the first. It was later confirmed that the targets were met in full; see Section 4.2.1 below.

SAF surveys

In addition to the outcome surveys linked to verification and payment, CPs also adopted a SAF to track progress in relation to four dimensions of sustainability: functionality; equity and inclusion; institutional and financial sustainability; and environmental sustainability. Functionality data were obtained from the CPs' own outcome surveys, but for other aspects they appointed consultants to undertake additional studies (one for each theme) with WEDC advising on design. CPs were disappointed with the quality of the surveys and analysis in the first round, though some useful findings were generated which triggered a management response – for example, providing training for social organisers on gender and equity. There were two rounds of SAFs planned: one towards the end of the output phase and one at the end of the outcome phase. The initial rounds of SAFs were delayed until into the outcome phase, with the delay reportedly being related to capacity constraints and time pressures in the output phase. The sequencing of the SAF rounds and the development of the outcome-phase strategy suggests there was a missed opportunity to use the results of the initial SAFs to inform programming earlier – though it should be highlighted that an element of the SAFs was a management response and the later assessments did still trigger changes in programming.

Other monitoring activities

Apart from the surveys outlined above, CPs and their partners continued their internal monitoring practices, and conducted some additional surveys, as part of routine project management. WaterAid, for example, conducted three knowledge, attitude, and practice surveys, firstly at baseline in 2014 and then during the outcome phase at the end of 2016 and late 2017.

Personnel from the CCU, CPs, and IPs all made routine monitoring visits to project sites and the total number of visits was substantial – so many, in fact, that some households began to complain at the number of people coming to check on toilets which they had built themselves with their own funds. The extent of field monitoring by consortium partners and IPs was not a specific requirement of the MVOC but rather part of the programme's approach to managing the risks associated with unexpectedly poor results – and therefore payment.

3.5 SAWRP in Bangladesh

3.5.1 Country context

Progress in WASH

Bangladesh's population has experienced a 20% gain in water access and 29% in sanitation access since 1990 and today close to 98% of Bangladeshis drink from an improved water source, while 63% uses an improved toilet. However, rural water supplies and sanitation remain seriously inadequate due to the poor quality of service that most people receive. Though most households have a plentiful and accessible supply of water, surface water sources are often contaminated (chemically or biologically) while in many locations groundwater is prone to problems of salinity or high levels of naturally occurring arsenic. Regarding hygiene, just 28% of the population have access to handwashing facilities with both water and soap, and infant faeces are often not

disposed of safely. Furthermore, only 10% of the population report using an appropriate water treatment method in their household[20].

Whilst access to improved water sources in rural areas is comparable across wealth quintiles, the poorest still have significantly lower access to improved sanitation than the rest of the population.

Institutional context

Unclear and overlapping allocations of functions, funds, and functionaries are a binding limitation to improving WASH services in Bangladesh and there has been only limited devolution to lower tiers of governance. Though assigned the role of service provider, many local government institutions do not have the technical or financial capacity to deliver and sustain high-quality WASH services. Furthermore, the current environment does not adequately incentivise private sector participation[21].

For SAWRP, an immediate institutional challenge was the frequent rotation of key government staff at all levels. This affected continuity as orientation had to be repeated and new relationships established each time a post holder changed.

Socioeconomic and cultural context

Bangladesh experienced a political crisis in 2015 that included a long strike by government staff. This was a significant obstacle because both WaterAid and Plan International worked in close collaboration with government agencies, especially at *upazila* (sub-district) level where government has a formal role in the coordination and monitoring of WASH activities and in ODF verification.

Physical and environmental context

Seasonal flooding is common in many parts of Bangladesh. This complicates the safe management and disposal of human waste and can cause serious damage to both toilets and water supply services. A high population density in many areas adds to the associated public health risks. The partners in Bangladesh intentionally targeted communities in flood-prone '*haor*' (wetland) areas.

3.5.2 Scope and scale

Including the 'School of Five' component implemented by Unilever, SAWRP worked in a total of 33 districts.[22] Of these, the community sanitation and hygiene component led by Plan International and WaterAid focused on 10 districts; see Table 2. Plan International aimed to achieve *upazila*-wide access to toilets using CLTS plus some elements of sanitation marketing, working with local masons (entrepreneurs).

While Plan International generally avoided hardware subsidies, at the programme start they introduced a special subsidy for the poorest to encourage and enable them to install (or upgrade

[20] This paragraph is adapted from World Bank (2018) *Promising Progress, A Diagnostic of Water Supply, Sanitation, Hygiene and Poverty in Bangladesh.* https://openknowledge.worldbank.org/handle/10986/29450
[21] This paragraph is adapted from World Bank (2018) *Promising Progress, A Diagnostic of Water Supply, Sanitation, Hygiene and Poverty in Bangladesh.* https://openknowledge.worldbank.org/handle/10986/29450
[22] This refers to the programme as it operated from inception and does not take into account the additional seven districts taken up under the programme extension, which falls outside the scope of this evaluation.

to) improved latrines with durable superstructures. Amongst other things, the intention was that, when the poorest households installed improved facilities, others would notice and feel motivated to do the same. Subsidy beneficiary households were identified through a participatory process at the planning stage in each targeted community. Plan International would offer to pay a share of the cost of latrine installation (Plan Bangladesh bore the material cost and installation of the sub-structure, beneficiary borne material cost and installation of the superstructure); and in some cases other community members reportedly also contributed. Entrepreneurs (artisans) participating in the sanitation marketing component would then install the facilities at a typical cost of BDT 4,000 to 5,000 ($48–$60) including superstructure. In some cases, they offered households payment in instalments.

3.5.3 Suppliers and IPs

The lead suppliers already had substantial CLTS experience when the programme began and built on that, applying established good practices. SAWRP was not, however, scaling up a specific existing programme.

WaterAid worked through local IPs while Plan International deployed their own implementation teams which worked closely with local government, signing an implementation memorandum of understanding with each participating UP. They worked in a total of 81 UPs of eight *upazilas* in the six districts listed in Table 2. In each UP there were roughly 15,000 to 20,000 households; a total of around 100,000 people. They also had other ongoing (non-WASH) programmes in five of the *upazilas*. Of the other three, one was coastal (more geographically challenging than other project locations, and with a larger proportion of marginalised people), while two were in the *haor* areas and thus subject to regular flooding.

3.5.4 Strategy changes between programme phases

As in Pakistan, the Bangladesh country programme did not map out an outcome-phase strategy until late in the output phase, though significant reductions in both field staff and budgets were anticipated for the period beyond December 2015. When the output phase ended, suppliers reduced the number of field staff drastically in line with their smaller operational budget.

Plan International also supported water quality monitoring. This included the development of community-level water safety plans and the provision of simple basic kits for communities, which offered a simple 'yes/no' result on the portability of water in terms of bacterial contamination.

In 2016, Plan International Bangladesh also adopted an 'ideal mother' initiative which involved working with groups of young mothers. This was unrelated to a similar initiative introduced by WaterAid in Pakistan (indeed, management were not very aware of it).

3.5.5 Approach to sustainability

During the outcome phase, WaterAid introduced some new initiatives on sustainability. This included making links with microfinance institutions to encourage and enable people to upgrade their toilets, for example by building a better superstructure or installing a concrete slab. BCC was

also a major part of the outcome-phase strategy, with a focus on handwashing promotion. The programme encouraged people to install handwashing devices such as tippy taps or containers with taps.

Plan International also focused more on the software side during the outcome phase. Community engagement and behaviour change remained a priority, but there was also more engagement with government functionaries, including efforts to make UP WASH and ward-level committees functional. These committees had a role in supporting sustainability, for example by making sure that, when latrines filled, households emptied or replaced them and continued to use them.

In addition to these activities, Plan International laid emphasis on hygiene promotion in schools. This was separate from the School of Five component led by Unilever and targeted different schools – some 800 in total. This included the provision of WASH facilities in some of them, though this was additional to the results for which Plan International was paid under PbR. These results were not reported to DFID and remained 'invisible.' There was also some installation of communal WASH facilities in densely populated communities, although, again, these were not part of the numbers reported or claimed to DFID.

3.5.6 Supplier monitoring approach and verification

The MV framework for the outcome phase in Bangladesh was much the same as in Pakistan, with Ipsos MORI conducting surveys to measure progress against the handwashing practice indicator while Plan International and WaterAid led surveys on the water supply and sanitation use indicators.

In the case of WaterAid and Plan International, Ipsos MORI provided initial training and then WaterAid and Plan International's partner NGOs conducted the water and sanitation surveys with WaterAid, Plan International and the CCU - providing supportive supervision. As in Pakistan, enumerators did not survey in their own implementation upazilas but swapped with other staff. Following Ipsos MORI training their own staff managed the surveys, using staff from the targeted UPs as enumerators. For 2nd and third outcome surveys, CCU Bangladesh provided the training to the selected enumerators and also provided supportive supervision.

As well as the surveys, WaterAid Bangladesh also tracked all e-Pact indicators via six-monthly sampling. To help improve data collection and management, during the outcome phase WaterAid Bangladesh introduced a live online data management system known as 'mPMIS' WaterAid continued to use their proprietary 'mWater' system in addition to WaterAid's version of 'mPMIS' live online data management system for tracking outcome progress. During output phase, WaterAid introduced 'mPMIS' to tract and record PbR indicators from the very beginning of the programme. Plan International adopted same type of 'mPMIS' of mobile phone-based data collection and reporting, though it took some months to get this running smoothly.' Amongst other things, this involved uploading photos of water and sanitation facilities installed or improved under the programme. Like WaterAid, Plan International adopted a system of mobile phone-based data collection and reporting, though it took some months to get this running smoothly.

The first outcome survey was the subject of some confusion at field level, but the challenges were ironed out and subsequent surveys were less problematic.

Apart from the outcome surveys linked to PbR payments, the country partners also commissioned SAF surveys as in Pakistan, investigating the following dimensions of sustainability: functionality; equity and inclusion; environment; and institutional and financial. This included some water quality testing, not only for bacterial contamination but also for arsenic, iron, and fluoride.

3.6 Linkages or synergies with other WASH programmes

In both SAWRP countries, other WASH initiatives were also being implemented, which created a more favourable enabling environment within which the WASH Results Programme operated (see table below). For example for SAWRP in Pakistan, activities were aligned with Pakistan's *de facto* national sanitation strategy, but the government was not directly involved in implementation.

Notwithstanding these synergies, it does not unlikely that there was direct overlap between the other WASH initiatives and the WASH Results Programme being implemented in the same communities. It was a requirement for suppliers to demonstrate to DFID that they were the only significant actors implementing WASH activities in their locality.

Table 15: Overlaps and synergies with other WASH programmes

Country	Donor	Programme	Timeframe
Bangladesh	DFID	Accelerating Sanitation and Water for All (ASWA)	2013-2019 (Phase 1)
	World Bank	OBA sanitation microfinance program	2016-2017
Pakistan	DFID	Accelerating Sanitation and Water for All (ASWA)	2013-2019 (Phase 1)
	Implemented by UNICEF	Sanitation Programme at Scale in Pakistan (SPSP Rural)'	2013-2017

Source: The databases that were searched to identify WASH synergies included DevTracker [DFID], World Bank Projects & Operations, African Development Bank Project Portfolio, Asian Development Bank Projects, WSSCC Global Sanitation Fund Countries and the UNICEF Evaluation database

4 Findings and analysis

4.1 Relevance

Box 5: Overall evaluation questions related to this section's discussion

Detailed evaluation question (DEQ) 1.1: To what extent were the programme objectives clearly articulated?

DEQ 1.2: To what extent does the programme's design (i.e. the TOC) set out a clear and realistic process for how programme activities will achieve the intended outputs, outcomes, and impacts?

DEQ 1.3: To what extent were the scale and pace of the programme (including the December 2015 deadline) realistic for achieving intended outputs and outcomes given the capacity of suppliers and their local partners?

DEQ 1.4: To what extent was the PbR modality appropriate for achieving sustainable and inclusive WASH outcomes, given the capacity of suppliers and the timeline of the programme?

DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?

DEQ 1.6: How likely was it that the programme design would encourage suppliers to propose 'innovative WASH interventions'?

DEQ 1.7: How likely was it that the programme design would encourage inclusive outputs and outcomes?

DEQ 1.8: How appropriate was the WASH Results Programme's design for achieving the programme 'learning objectives'?

DEQ 1.9. To what extent was the design of each consortium sub-programme appropriate for achieving DFID's key objectives?

4.1.1 WASH context and alignment with national policy context

In Pakistan, the programme works within the established PATS framework. Indeed, Plan International was directly involved in the development of PATS via the UNICEF-led RuSFAD beginning in 2010. SAWRP also seeks to support sector decentralisation by supporting and encouraging the role of local government (especially UCs) in supporting the sustainable use and maintenance of latrines post-ODF certification.

In Bangladesh, government commitment to CLTS has waned somewhat in recent years, partly because it was closely associated with a previous administration but also because open defecation has already been reduced to less than 1% and the principal challenge now lies in achieving safely managed sanitation as per the SDGs. Similarly, with water supply, the challenge is not so much access as improving the quality and level of service. SAWRP is supporting the shift in focus towards the SDGs by including an element of sanitation marketing that helps low-income households upgrade to durable, hygienic latrines.

4.1.2 Programme design

Implementation approaches: The programme objectives were relevant in that they were closely aligned with DFID's design and priorities. In Pakistan, much of the work undertaken by SAWRP

represented the extension, expansion, or replication of recent or ongoing projects managed by consortium partners, building upon operational approaches and partnerships that were tested and improved under these initiatives. The consortium was therefore in a good position to deliver its expected results at scale. Pre-existing relationships with several national IPs also contributed to a favourable programme environment in Pakistan.

Senior managers highlighted that 'the PbR modality in itself was not an important factor in the decision to engage or not with the WASH Results Programme'. For them the payment modality was a secondary issue – it was 'a means to an end'. In Bangladesh, both partners had previously worked on WASH and in some of the areas included in the programme. Broadly speaking, the approaches used were not new to the WASH sector, however.

Output/outcome phase split: Neither WaterAid nor Plan International had routinely included a dedicated outcome phase in their earlier WASH projects in Pakistan and Bangladesh, and there was no clear roadmap for delivering the outcomes when SAWRP began. Nevertheless, programme managers considered it very positive that SAWRP had an additional two years to focus on consolidating behaviour change and fostering sustainability. As highlighted above there was a large reduction in staff strength in Bangladesh between the output and outcome phases, although there was nevertheless much further work to do, especially on the software side (e.g. strengthening local government ownership and providing further training for communities). WaterAid felt it necessary to continue quarterly beneficiary engagement, but now had insufficient frontline workers to do it. Having had 140 during the output phase, they now had only 48 and the project engineer post was also cancelled. With reduced personnel, existing staff reportedly had to take on multiple tasks. This scenario was almost same for Plan International Bangladesh.

In both countries there were numerous examples of suppliers undertaking work not linked to payment or reflected in the results. Such aspects are considered 'unintended' as they were not specifically a programme goal. Examples include:

- Plan International in Bangladesh provided some additional facilities that were not reported covered by payments under the PbR modality. These included community toilet blocks in some densely populated communities with limited space for household facilities and rainwater-harvesting facilities where ground water supplies were unavailable or satisfactory.
- Both Plan International and WaterAid offered community-level 'incentives' for the achievement of ODF status. These included additional water points in the community, water filtration systems (slow sand filters) for existing systems, and communal septic tanks.

In general, the implementation approaches used and programme areas targeted were not significantly influenced by the PbR modality, though the division into two programme phases did significantly affect the supplier's approach. Most programme logframes or results frameworks anticipate the delivery of both outputs and outcomes, but having a dedicated outcome phase is rare and its inclusion was felt by many respondents to be a significant and positive feature (although this positivity is balanced against criticism from partners of the tight timelines of the output phase). Furthermore, being held accountable for delivering against the outcome indicators helped to ensure that suppliers remained active in community support and engagement to facilitate the transition from outputs to outcomes up to the programme end. Having said this, it could also be argued that enhanced accountability for outcomes could also be incentivised under other contracting modalities.

Strength of evidence: moderate

Clear and consistent feedback as to which design features significantly affected implementation, and the relationship between these choices and the PbR modality.

4.1.3 Coordination and synergies with other initiatives

Plan International Pakistan had a large DFAT-funded sanitation programme operating in parallel with SAWRP. This was grant-funded and helped to ensure the smooth funding at country level to cover general operational overheads. WaterAid core funding was used to fund School WASH interventions in Pakistan, complementing CLTS.

In Pakistan two other large, multi-donor WASH programmes were under development in Sindh at the time of the evaluation mission. Both had a strong sanitation and hygiene component, but they were not yet fully operational at the time of the mission. Plan International Pakistan's new strategy entails ceasing WASH programmes in Pakistan so it is unlikely it will have future involvement in the programme in Sindh, although many of the local IPs may well be involved in the MSAN.

In Bangladesh, Plan International was an IP under the UNICEF-led ASWA programme, which was also funded by DFID and had similar objectives. This ran from 2013 to 2016 and its contribution operated in 26 UPs across four districts, with a budget of $1.7 million. In addition, Plan had a small 'WASH in Schools' project funded by private grants. Nonetheless, SAWRP was by far the biggest programme for the suppliers over this period, with a budget of over $15 million.

In Pakistan the interviews with programme managers suggest that the programme was largely implemented independently of other WASH programmes, albeit sufficiently well-coordinated with government approaches and partners. In Pakistan the team did not identify significant missed opportunities for greater coordination. In each of the three provinces the programme worked in in Pakistan there was a government-led WASH programme, the functionaries of which were engaged with as part of SAWRP implementation. Furthermore, provincial WASH staff were positive surrounding the coordination efforts of the partners and did not identify serious issues.

Strength of evidence: suggestive

Other programmes identified but it remains unclear if opportunities were missed for synergies at the field level.

4.1.4 Timelines and how realistic the targets were

During the midline case study in Pakistan it was the suppliers' perception that DFID was pushing them to achieve MDG targets rather than focusing on the sustainability elements of the programme. This focused attention on results but put pressure on programme staff. Consortium partners and IPs estimated that at least six months were needed to implement the CLTS approach in any given community. However, IPs indicated that in exceptional circumstances, such as working in a new community or working in areas where 'direct incentives' had been the 'norm' in past WASH interventions, then achieving higher community coverage levels took longer – between

six and 18 months. In cases where ensuring community engagement was difficult, then IPs sought political support through the UC Chairmen or through religious leaders (in mosques). At midline, the LGD in Sindh remarked: *'[The] SAWRP outcome phase is going to be more crucial and challenging as it will require more engagement with all stakeholders to lead the process towards making the interventions sustainable, especially in relation to the behavioural change component'.*

The verified results confirm that the programme targets were indeed achievable; in fact, the level of overachievement indicates that the programme could have been even more ambitious, stress permitting. In both countries, partners highlighted that the overachievement was planned for as a risk-management strategy from the start. Unilever Pakistan were quite explicit that with hindsight they would have been more ambitious; they were reportedly conservative in setting their targets.

Strength of evidence: moderate/strong

The impact of the timelines on implementation was significant and something the evaluation team received clear and consistent information on.

4.1.5 Operationalisation of the PbR modality

Plan International UK were the contract holder with DFID, and between the three lead partners (Plan International UK, WaterAid UK, and Unilever) there was a PbR arrangement where partners were paid *pro rata* against achievement in line with a fixed price per beneficiary[23]. For Plan International and WaterAid, the flow of funds to the country offices was via the UK offices,[24] which also had a role in interfacing with DFID. IPs were contracted using grants. In the case of WaterAid in Pakistan and Bangladesh, the UK office explicitly underwrote the country office – that is, should the results not be achieved in any country the country office would still receive the full payment and the financial loss would be absorbed by the UK office using other unrestricted funding and reserves. With the exception of WaterAid in Bangladesh (see Section 4.3.1) the PbR risks were not passed onto local IPs.

At the design stage a risk-management strategy was developed that included a 'risk premium' charged to DFID for the organisational risk associated with PbR. SAWRP had a contingency fund (4%) that could be used by countries to cover eventualities such as natural disasters (e.g. flooding). Requests are made by IPs to the CCU, then forwarded to the Executive Team in London, where allocation was decided at the monthly coordination meetings. The fund was reportedly needed in Pakistan due to the sharp decline in the sterling exchange rate following the Brexit referendum in June 2016. This was also applicable for Bangladesh.

Internal risk-sharing arrangements were not raised by suppliers as significant constraints to implementation. The consortium partners were all large organisations for which pre-financing was not a great concern, and PbR risks were not generally passed on to the IPs[25]. The risk-sharing arrangements in relation to efficiency are discussed further in Section 4.3.1.

[23] The evaluation team understand that this varied by partner and country and differed by the price per beneficiary in the main contract with DFID, though we lack details of the internal contracting arrangements within SAWRP.
[24] In some cases with the head office assuming the PbR associated risks as opposed to the country offices (see Section 4.3.1).
[25] WaterAid Bangladesh did pass on some risks, and there was a pre-financing element for IPs. There was also more of a focus on upside incentive payments for overachievement rather than downside penalties for underachievement.

Strength of evidence: moderate

Contracting arrangements within the consortia are clear from interviews, but the evaluation team lacked access to the specific grant/partner agreements/contracts used within the consortia.

4.2 Effectiveness

This section considers the relationship between the outputs and outcomes in the programme. The WASH Results Programme was unusual compared to other WASH programmes of the time in that there were explicit payments linked to outcomes combined with intensive monitoring of outcomes linked to the verification. At the time of the WASH Results Programme's design the WASH sector globally lacked reliable benchmarks for outcome-level achievements and this to some extent explains why the three suppliers negotiated different levels of ambition at outcome level, as well as in some cases different outcome levels for different countries within a consortium.

Box 6: Overall evaluation questions related to this section's discussion

DEQ 2.1: Did the programme achieve the intended outputs at scale?

DEQ 2.2: To what extent have the utilisation of water and sanitation services and the uptake of hygiene practices reached all members of target populations (inclusive outcomes)?

DEQ 2.3: To what extent have services continued to function and have behaviours continued to be used since their initial implementation (sustainable outcomes)?

DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?

DEQ 2.5: Under which circumstances did the PbR framework help/hinder the achievement of intended outputs and outcomes?

DEQ 2.6: Under which circumstances did the PbR framework affect the quality of programme implementation (positive or negative)?

DEQ 2.7: Under which circumstances did suppliers implement innovative approaches and focus on learning?

4.2.1 Effectiveness by programme area

In both Bangladesh and Pakistan there were programme elements that were not well captured in the results that had benefits for the communities – examples include incentives (such as waterpoints, water filtration systems or community sanitation schemes) for achieving ODF status in Pakistan and the construction of communal facilities in Bangladesh. These are discussed more in Section 4.4.3 below.

In Pakistan and Bangladesh the first rounds of outcome surveys found that both the water supply and sanitation use targets were met in full, with considerable overachievement in cases. There was, however, a slight shortfall in the case of handwashing practices at points in both countries. There was also a negligible difference in outcome-level achievement between the two countries.

4.2.1.1 Sanitation outcomes

In both countries the results were well above the target and improved across survey rounds. In the case of SAWRP the sample frame for the endline surveys in sanitation was the output-phase beneficiary database, meaning that the endline surveys are a panel survey of the output-phase beneficiaries. This was a different approach to that taken by either SWIFT or SSH4A, which surveyed the entire population in their implementation areas (including beneficiaries and non-beneficiaries). As such, the SAWRP results provide a view as to the extent to which specific output-phase beneficiaries continued to use their facilities but do not give an indication of the population-wide usage levels in programme communities.

An important aspect of interpreting the outcome data is the indicators used. In the case of SAWRP, a RAG rating was used to measure use. For sanitation, an Amber or Green rating for self-reported use coupled with a Green rating for observed functionality would result in the results target being met being met and correspond to payment. The percentages below represent the overall proportion of households for which there were only a Green or Amber ratings (i.e. no Red). This includes cases where the self-reported use was Amber ('Would you say that most household members use this latrine when they are at home?' = 'Most of the time' *or* 'Don't know', which was rated Amber).

Table 16: Verified outcome-phase results of SAWRP in sanitation

Area	Indicator and target	Pakistan		Bangladesh		
		Q1 2017	Q4 2017	Q4 2016	Q2 2017	Q4 2017
Sanitation	75% of poor people across project districts continue to use basic or improved latrines	96.1%	99%	99%	99.1%	99.2%

Source: Initial submissions of Q1 2017, Q2 2017, and Q42017 verification reports

In both Pakistan and Bangladesh the programme remained active at local government and community level throughout the outcome phase, despite staff reductions. The second outcome survey report noted two particular features of programme strategy that contributed to the final outcome. In Bangladesh, local government institutions and the country partners made quarterly joint monitoring visits to consolidate the use and maintenance of WASH facilities and promote handwashing, while programme monitoring systems included mechanisms to ensure that timely corrective action was taken to address any shortcomings or challenges identified. Plan International Bangladesh also promoted 'sanitary marts' during the outcome phase in most of the working communities. WaterAid in Bangladesh continued regular monitoring visits on quarterly basis during the outcome phase. In addition, WaterAid in Bangladesh introduced a microfinance component in the outcome phase to help households upgrade from basic latrines built in response to CLTS triggering. Both country partners included this feature in the programme extension. Similarly, Plan International in Pakistan focused on upgrading latrines from 'basic' to 'JMP improved' over the course of the outcome phase. The inclusion of sanitary marts was cited as an enabling factor in programme design here. Additionally, Plan International Pakistan also provided finance on very easy terms for sanitary mart operators and microfinance was also available to households in several districts via NRSP. In Lodhran, LPP helped to establish mobile sanitary marts.

Though not reflected in the results data, suppliers in Pakistan reported an ODF conversion rate of around 98% (note this has not been verified by e-Pact). By global standards, this is exceptional; Plan International's own global research has found that CLTS programmes rarely achieve more than 60%. Interestingly, the bulk of ODF certification happened during the outcome phase. Suppliers explained that most of the communities actually became ODF during the output phase but it took a long time for government agencies to complete the formal certification process given the number of communities involved. Plan International Pakistan also reported that 94% of the basic latrines constructed during the output phase had been converted to improved facilities by the time of the final outcome survey.

Programme stakeholders were asked how they accounted for such a high ODF conversion rate. Several put it down to concerted and prolonged promotional efforts, which some attributed to the PbR modality and more specifically the associated accountability for outcome targets.

However, it should be noted that the ODF verification protocol is not standardised across provinces in Pakistan and there are no official national ODF criteria, hence there is some variation in the criteria adopted by different implementing agencies and provinces. In the case of some of the certification in Sindh, many of the certified communities were village 'clusters' rather than entire revenue villages (which typically contain up to 10–12 communities) and ODF definitions and certification standards are not consistently applied across or within districts. Notwithstanding these caveats, the results are very positive and some additional, unexpected positive outcomes were also reported by country partners:

- for the first time, communities in Thatta District began building toilets in their seasonal fish camps as well as their villages; and
- in Badin District, a number of households built toilets attached to their home for first time.

Environmental factors were very significant in Bangladesh as flooding caused substantial infrastructure damage in some programme areas, leading to people migrating and often taking their latrine slab with them. A lot of repairs were done too, however, and WaterAid Bangladesh used programme contingency funds in one *upazila* to support rebuilding of latrine sub-structures while microfinance institutions provided loans for superstructures. In areas which did not experience such serious impacts, seasonal floods nevertheless complicated local travel by programme staff and residents, as well as the transporting of materials.

The verified outcome results suggest that the programme was effective in maintaining latrine functionality and that latrines were used, although it should be noted that the indicators used mean that a degree of non-use could be present. Programme activities are seen to have contributed to the achievement of outcomes, and it is the evaluation team's view that without the outcome phase functionality would likely have been lower.

Strength of evidence: moderate

Analysis is based primarily on the verified survey data, and complemented with interviews with supplier staff. The verified data is regarded as credible, albeit with the caveat that no third-party data was collected against which these data were triangulated.

4.2.1.2 Hygiene outcomes

For SAWRP the original outcome target was expressed in terms of beneficiary numbers and implied that 90% of those reached with hygiene promotion messages would continue to wash their hands with soap. This target was later revised as part of the MVOC development and a composite indicator based around three components (knowledge, reported behaviour, and observed behaviour) was developed with pp target increases. It is also worth noting that the targets were slightly different between Bangladesh and Pakistan and partners – with Bangladesh having a slightly higher target for the pp increase in self-reported behaviour. A final significant point to consider on measurement is that in Pakistan all sanitation beneficiaries were assumed to also be hygiene beneficiaries, and this was the basis on which hygiene beneficiaries were calculated and reported to DFID. That is, if a household constructed a latrine during the output phase those household members were assumed to both be a sanitation beneficiary and reached by hygiene promotion.

For Plan International and WaterAid the target was met in most cases, though in both countries there were points where the target was not met on the knowledge component. In Pakistan there was considerable debate surrounding the appropriateness of this indicator as men were interviewed and two of the critical times (related to childcare and food preparation) were not seen as relevant to them. There was also further debate surrounding the methodology. In the Q1 2017 verification round, Ipsos MORI used a random walk methodology and consequently interviewed both non-beneficiaries and beneficiaries as part of the data collection. This was seen by SAWRP country management in Pakistan to have negatively impacted the results through including non-beneficiaries. Moreover, in Bangladesh country programme management highlighted that the sample may have even included areas in which they had not worked. The lower-than-target achievement resulted in a small payment disallowance – Section 4.3.1 discusses how this was managed between partners. The issues with measurement were taken into account by DFID when deciding on the payment amount and subsequently the outcome survey methodology was adjusted such that only beneficiaries were interviewed.

Table 17: Verified outcome-phase results of SAWRP in hygiene (household)

Area	Indicator and target	Pakistan		Bangladesh	
		Q1 2017	Q4 2017	Q4 2016	Q4 2017
Hygiene	Reported practice: Percentage of respondents saying 'before eating' increases by: 15pp(B)/10pp(P)	+9pp (81%)	+33pp	+32pp (77%)	+38pp
	Observation: At least 10pp (B&P) more of these interviewer observations should reveal the presence of soap or a soap substitute, compared with baseline.	+7pp (74%)	+27pp	+27pp (78%)	+32pp
	Knowledge: The percentage of people who can name three or more critical times for handwashing increases by 15pp (B&P) compared to the baseline.	+2pp (37%)	+11pp	+9pp (26%)	+20pp

Source: Initial submissions of Q1 2017, Q2 2017, and Q42017 verification reports

Table 18 presents the results of the 'child survey', which focused on the hygiene outcomes under the programme component implemented by Unilever. As with the survey of those reached by household-level hygiene promotion there were issues with the specification of this indicator. Specifically, in Pakistan the midline and endline results on the knowledge indicator were lower than at baseline – which must be set in the context of higher results for reported practice and observed behaviour. As with the household-level results this highlights the complexity of capturing hygiene behaviour change in a meaningful way and the complex relationship between the knowledge component and the other indicator components. In the case of Pakistan the negative progress on knowledge was attributed to there being a high baseline figure for the 'household survey' – with country programme management feeling that this high baseline results may be related to sample and non-sample error. Global programme M&E staff additionally highlighted that seasonality was also identified as a factor. It should also be noted that due to attrition[26] the baseline figures were adjusted between the midline and endline surveys.

Given the uncertainties surrounding the measurement – and with the exception of the knowledge indicator in Pakistan – the results are overwhelmingly positive, with the targets largely achieved and considerable overachievement. In both Pakistan and Bangladesh there was also improvement on the self-reported behaviour indicator between surveys and the results were high – in both countries around 80%–90%.

Table 18: Verified outcome-phase results of SAWRP in hygiene (child)

Area	Indicator and target	Pakistan*		Bangladesh	
		Midline	Endline	Midline	Endline
Hygiene (Child)	**Knowledge:** Percentage of respondents able to state all the five times increases by 10pp	-39% (15%)	-27pp (23%)	+16pp (62%)	+33pp (79%)
	Reported practice: Percentage of respondents saying 'everyday'(P) /'everyday/most days/some days' (B) increases by 10pp	+19pp (74%)	+20pp (80%)	+28pp (73%)	+43pp (88%)
	Observation: For children practising fewer than six steps of quality handwashing at Baseline (i.e. or less), the mean number of quality handwashing steps they undertake at mid/endline will increase by +1 step	+2.3 steps (5.7)	+3.0 steps (6.6)	+3.1 steps (6.9)	+4.0 steps (7.8)

Source: Q4 2016 verification report

*Note: The baselines for the mid and endline survey analysis in Pakistan are different because they reflect attrition rates during the surveys. The child surveys are longitudinal cohort studies and the sample size during the mid and endline was reduced if the same children could not be found.

Across counties and programme components the targets for hygiene were largely achieved – though with the notable exception of Pakistan in Q1 2017[27] and on the knowledge component in Pakistan in both rounds and Bangladesh at midline. Compared to the other programmes, a comment as to how effective these programmes could have been is extremely hard to judge as there are few benchmarks for what is a reasonable level of achievement in different contexts. One benefit is that the data do provide such a benchmark for future

[26] Children moving out of the schools sampled.
[27] Partially related to measurement issues.

programmes, although, as noted throughout this section, measurement related to hygiene is challenging and issues with measurement and changes to the methodology across the programme make these data more difficult to use for comparing performance across the programme.

Strength of evidence: moderate

As with sanitation, with the addition that the measurement issues related to hygiene and the lack of reasonable benchmarks in this area hamper the analysis of effectiveness.

4.2.1.3 Water outcomes

As with sanitation the water results are overwhelmingly positive, with significant overachievement against targets in both countries. Again as with sanitation, the results presented to DFID for payment arose from an aggregation of RAG ratings. For a water point to be ranked (functional) it would have to be judged by the enumerator to be an improved water source functional at the time of the visit (i.e. water could be drawn) and that a sample of users interviewed reported that they could draw water from an improved source throughout the year.

In terms of the factors that supported functionality, as with sanitation it was emphasised by programme management that accountability for results was a strong influencing factor in maintaining functionality. That said, it should be noted that water was a much smaller component of the programme with regards to beneficiary numbers than sanitation or hygiene. Additionally, some water supplies improved under the programme in Bangladesh were affected by high iron concentrations, salinity, and dropping water tables – this added to the cost and challenge of providing and maintaining adequate, safe supplies.

The outcome results support the view that the programme was effective in ensuring continued functionality of systems until the end of the programme. However, it is noted that this is separate from discussions surrounding the prospects for sustainability (addressed in Section 4.5.1) and that partners played an active role in ensuring functionality over the course of the outcome phase.

Table 19: Verified outcome-phase results of SAWRP in water

Area	Indicator and target	Pakistan		Bangladesh		
		Q1 2017	Q4 2017	Q4 2016	Q2 2017	Q4 2017
Water	90% of poor people across the project districts continue to use reliable, safe drinking water sources	100%	100%	97.1%	99.2%	98.6%

Source: Initial submissions of Q1 2017, Q2 2017, and Q42017 verification reports

Strength of evidence: moderate

As with sanitation.

4.2.2 Quality of results

The evaluation found no strong evidence from the outcome phase of PbR effects on the quality of programme results, whether physical or in terms of behaviour change, or the capacity and motivation of programme beneficiaries or local government institutions. However, some respondents did note that accountability for outcome-level results meant that they had to follow up and ensure that the functionality and use of WASH facilities continued until the end of the outcome phase.

Country managers in Bangladesh also highlighted that during the output phase they were required to deliver to a specified quality of result, though there was some inconsistent feedback from country programme management from the lead organisations on this point. Some emphasised that the PbR modality did incentivise a focus on quality, but others emphasised that these concerns were already operational priorities for the suppliers before the introduction of a PbR framework. Thus, some respondents did not consider the quality of results under SAWRP as noticeably better than what was a typical non-PbR approach.

At the midline assessment in Pakistan there was agreement amongst all those interviewed at field level (IP staff) that quality of implementation was better under PbR than in grant-based programmes due to intense mobilisation, with the monitoring requirements adding a layer of scrutiny of the results that acted to incentivise quality[28]. Nonetheless, it was highlighted by partners that the pressure to deliver ambitious targets by December 2015 meant that some processes were rather rushed (one IP commented that it was like an emergency response project). Plan International Pakistan, for example, simplified its CLTS pre-triggering and triggering processes and country staff described the community action plans developed as somewhat rudimentary. However, the outcome phase gave them space to 'catch up' and consolidate community commitments.

Different suppliers in different countries aimed for different standards of latrine. In Bangladesh both suppliers aimed to construct 'improved' facilities from the outset; similarly, WaterAid in Pakistan had a focus on constructing improved facilities from the outset. Plan in Pakistan, however, accepted a lower standard of latrine in the output phase with a view to securing the improvement of these facilities in the outcome phase. Plan International Pakistan reported to the evaluation team a high conversion rate from 'basic' to 'improved' facilities in the outcome phase, although these results were not verified by e-Pact.

Overall, the evaluation is limited in the extent to which we can comment on the quality of results – though aspects of programme design are seen both to have supported greater quality as well as potentially harmed a focus on quality. There was also mixed evidence from interviews as to the effect of PbR on quality. The accountability for outcomes and the standards embedded within the indicators was seen by some to have supported a greater focus on quality. At the same time, however, the rushed implementation timelines during the output phase clearly placed pressure on staff – potentially undermining the quality of implementation.

[28] For example, LPP in Pakistan identified cases where latrines were rejected from the latrine database on the grounds of quality concerns.

Strength of evidence: suggestive

Analysis based primarily on interview evidence without any third-party data on facility quality. Additionally, there was some disagreement on key points among stakeholders.

4.2.3 Effectiveness of supplier monitoring systems

The discussion in this section centres on the degree to which the supplier monitoring systems generated data that was fit-for-purpose and the influence of the PbR modality on this. How efficient these systems were is considered separately in Section 4.3.3.

In both Pakistan and Bangladesh country partners were clear that while the monitoring and verification burden was heavy throughout – and exceeded the demands typical under grant-funded programmes – it nonetheless contributed to them strengthening their monitoring systems. Respondents in Pakistan and Bangladesh also emphasised that the monitoring burden lessened over the course of the programme as partners became used to the systems. For the extension (SAWRP II), the consortium was able to develop clear monitoring and reporting guidelines from the start. Additionally, according to both district-level supplier teams and IPs in Pakistan, having the entire team involved in monitoring activities helped in capacity building the whole team: *'monitoring wasn't limited to a designated job description'*[29].

The requirement to measure outcomes triggered significant debate among the partners and the MVOC took nearly two and a half years to negotiate. However, this debate triggered scrutiny of both the indicator selection and measurement. For example, WaterAid Pakistan found that being forced to identify meaningful indicators and behaviour change targets was a positive aspect of the PbR modality, and informed us that it has since incorporated some good practices such as spot and back checks into other, grant-funded projects.

As is common in WASH programmes the measurement of beneficiaries at the output level relied on the use of multipliers[30]. Scrutiny of these multipliers resulted in 'caps' being introduced on service quality grounds; for example, in Q4 2014 a cap of 14 people[31] per latrine was established for SAWRP latrines. The need for a cap on beneficiary numbers per facility was highlighted in recommendations from the lead verifier, although the precise numbers to be used did not form part of the recommendations. The cap on the number of beneficiaries per latrine led suppliers to modify their implementation strategy in households with multiple families. Similar caps were established for water systems under a similar logic[32]. With regards to hygiene, in Q1 2015 the independent monitor suggested school enrolment figures should not be used for Unilever's handwashing promotion interventions in Pakistan as enrolment did not necessarily reflect attendance. As a result, the verified beneficiary numbers for Unilever school activities were reduced by an average of 12.5%. This led to a review of the So5 monitoring systems at the global level. Unilever now uses

[29] Note that the team did not interview field staff in Bangladesh so are unable to comment at his level of detail for the Bangladesh programme.
[30] The number of people assumed to be using a particular piece of infrastructure – used to calculate beneficiary numbers based on the delivery of outputs such as latrines or boreholes.
[31] Up to that point 151 latrines constructed by WaterAid claimed more than the maximum 14 beneficiaries, amounting to 4.9% of the total claim to date.
[32] In Q2 2015, a cap of 100 people/handpump in Pakistan was established for small locally manufactured Abyaar handpumps, and a cap of 20,032 people was established for Afridev handpumps. At a later stage in Q4 2015, a cap of 400 people/water filtration unit, referred to as a Biosand filter, was established.

a different system for calculating So5 beneficiaries, i.e. attendance figures on the day activities take place are now used.

In the early stages of the programme, verifiers identified a number of shortcomings and suppliers had to take remedial action to bring their monitoring and reporting systems up to a standard acceptable the verification team and to DFID. As well as appraising the quality of monitoring systems overall, at times the MV team had to conduct appraisals related to the measurement of results against a particular milestone. The country verifier made periodic field visits to confirm how monitoring systems were operating on the ground. This included both spot checks (where the verifier accompanied programme staff as they collected data in the field) and back checks (where the verifier took a sample of completed reports and repeated the data collection process).

In both countries there were issues related to the measurement of hygiene. Specifically, the baseline figures in Pakistan across the indicator components were seen to be unexpectedly high by programme staff – to the extent that some of the country programme management in Plan and WaterAid suspected measurement error (though this could not be corroborated with data). Furthermore, there were issues with respondent sampling during the survey for the Q1 2017 verification round, in that the sample included beneficiaries and non-beneficiaries.

During the outcome phase partners in both countries continued to monitor the functionality of infrastructure outside of the MV requirements, partially as a risk-management strategy. In Pakistan, for both Plan and WaterAid this took the form of regular follow-up visits at the household level. In Bangladesh, one new monitoring initiative during the outcome phase was that suppliers began making quarterly joint monitoring visits with local government personnel, and reportedly found these useful as a means of encouraging and supporting the achievement of sustainable outcomes – as was increased rigour in programme monitoring generally.

In both Pakistan and Bangladesh the surveys were implemented by field staff, who were rotated across implementation areas so they did not conduct the survey in the same area in which they normally worked. Even so, in Pakistan there was some speculation among programme stakeholders at the national level regarding a possible conflict of interest in allowing suppliers to conduct the outcome surveys themselves – effectively 'marking their own homework' – but both DFID and the verifiers accepted this arrangement in advance on the condition that the verifier quality assured the methodology. The technical capacity of project field staff to serve as enumerators was also questioned by some, though verifiers noted that they followed operational guidelines diligently, and this was not a uniform view among national level stakeholders – with some feeling the social organisers had a transferable skill set[33] and were well trained while others felt it was asking them to perform a function beyond their job descriptions.

The verification was 'systems' based, with limited third-party data collected to validate the data produced by the supplier monitoring systems. In Pakistan[34] at midline the country verifier and his assistant highlighted a number of interesting points. Their contract provided for only a limited number of days of work per verification round. Most of the in-country verification work involved reviewing supplier databases and the supporting evidence submitted by suppliers, and conducting the systems appraisals. Neither the country verifier at the time of the midline nor their assistant expressed much satisfaction with the nature of the work. It should be noted that between the

[33] Predominantly skills related to engaging with community members.
[34] The country verifier was not interviewed in Bangladesh.

output and outcome phases a new country verifier was hired in Pakistan and they echoed some of the frustrations voiced by their predecessors at midline.

The benefits of the verification in regard to strengthening monitoring systems was largely confined to the suppliers within the consortia, with a limited focus on strengthening government monitoring. This was largely as, by design, the data collected drew exclusively on the SAWRP partner's own monitoring systems. In Pakistan[35] the provincial government stakeholders interviewed from the three provinces in which SAWRP worked highlighted that, while government systems were not strengthened through the programme, there were nevertheless some benefits to the government as it received more credible data on sanitation improvements in the province.

It is possibly the most uncontentious aspect of this evaluation that the verification process contributed to suppliers strengthening their monitoring systems. There is near universal feedback – from both supplier and verifiers – that this was the case, while this view is supported by clear examples and documentary evidence in the form of the verification reports. However, in light of clear evidence of improvement, the limited third-party data collected as part of the WASH Results Programme or verification approach mean that an overall assessment of the veracity of the data is not possible. While this is not to suggest that the evaluation team have strong grounds to question these data, it remains the case that the monitoring data drew on a range of assumptions and due the nature of the data collection in sample surveys there are limitations as to the extent these can be 'verified' without third party data collection.

Strength of evidence: Strong

Consistent feedback from suppliers as to the benefit of the verification to strengthening monitoring, and documentary evidence related to monitoring system strengthening through the verification reports.

4.2.4 Flexibility and innovation in practice

Innovation in this context is taken to refer to innovations in programme approach – that is, where there was an application of novel approaches to overcome previous challenges. This framing of innovation around programme approach is rooted in the assumption, widely cited in the literature that PbR enables greater scope for innovation by removing donor requirements related to implementation approaches.

Similarly, the framing of flexibility here is also rooted in the programme context and the PbR nature of the contract, i.e. under PbR suppliers are seen to have greater autonomy over implementation activities and budget as these are not reporting requirements to the donor[36]. The team recognise that this is a specific framing of flexibility[37].

[35] Again, note that local government actors were not interviewed in Bangladesh.
[36] Supplanted by reporting only on results.
[37] Particularly that flexibility is often framed in the ability to respond to changes in context, and distinguished from adaptation (a change in knowledge of a context). Arguably these framings are partially covered by the framing used by the evaluation team, but these were not used in the framing of questions in interviews.

4.2.4.1 Innovation

The evaluation found little evidence that the PbR approach stimulated innovation, and tried and tested implementation approaches were used. This was a key consideration when the consortia were formed as it increased the predictability of results and was seen as an important factor for programme success.

Strength of evidence: moderate/strong

As a topic covered by the contribution analysis this was a focus of the interviews at all implementation levels. There were consistent findings across stakeholder groups, although the team note that there was comparatively limited engagement with the implementation-level actors in Bangladesh.

4.2.4.2 Flexibility

At programme management level in both countries it was highlighted that there was a greater degree of flexibility in managing the budget and changing activities. It was emphasised that the key changes were greater autonomy over the budget and the lack of a need to have changes to activities approved by the head office or donor. WaterAid Pakistan staff in particular emphasised that this led to a greater sense of empowerment as they could focus on results and move budget and change activities accordingly. In both countries, managers were generally very positive about PbR and the effect it had on the programme. However, it is notable that this positivity was stronger in the outcome phase than the output phase[38]. As discussed earlier, the tight timelines of the output phase significantly influenced implementation – arguably more than the additional degree of flexibility afforded by the PbR modality. In both countries it was highlighted that the hard timelines were an area of inflexibility vis-à-vis other grant programmes they implement.

The additional flexibility was not cascaded to the implementation level. In both countries the local IPs were under grant agreements and were required to do financial and activity reporting against a workplan in addition to complying with the results monitoring and verification requirements. IPs were aware that the partners had greater budget autonomy and could agree to changes in operational approach if they considered it appropriate. Despite working through logframes, Plan and WaterAid allowed IPs to adjust the targets or renegotiate grant agreements to increase capacity if needed. Plan in Bangladesh did not have IPs but worked more directly with *upazila* staff. There, a memorandum of understanding was signed with the government counterparts, with the *upazila* doing quarterly reporting (financial and progress) to Plan International Bangladesh – though it should be noted again that the government counterpart staff were not interviewed as part of the evaluation so the team are unable to reflect their experiences.

Overall, there was consistent feedback from the country programme management level that the PbR modality contributed to there being greater flexibility in budget management and choice of activities at the programme level, although in most cases this greater flexibility was not cascaded to the implementation level. In Pakistan the evaluation team had greater sight on field-level implementation, and partners were clear that this greater flexibility was intentionally not cascaded to the implementation level. Comparatively, in Bangladesh the

[38] Bangladesh country programme management changed for WaterAid between phases and the Plan country programme manager was not interviewed at midline.

evaluation team had less sight on this, having not interviewed IPs in Bangladesh, and the country management in Bangladesh also highlighted that this greater flexibility was not cascaded fully.

Strength of evidence: strong

As a topic covered by the contribution analysis this was a focus of the interviews at all implementation levels. With consistent findings across stakeholder groups. Though the team note that there was comparatively limited engagement with the implementation level actors in Bangladesh.

4.2.5 Experiences of adaptation and learning at the implementation level

This section discusses the degree to which adjustments at the implementation level were made to improve effectiveness, based on insights gained from near real-time monitoring data or evaluations. This is distinguished from innovation in programme approach more broadly as well as from flexibility (as framed above). As with innovation and flexibility, this framing is rooted in the context of this programme and an assumption[39] that the PbR modality would incentivise such adaptation at the implementation level to ensure targets are met effectively and therefore facilitate or increase payments.

There was not strong evidence[40] that the PbR modality encouraged results-oriented problem solving. This is associated with the fact that partners were generally working under grant agreements with logframes and that the increased degree of flexibility was largely at the country programme management level as opposed to the implementation level (see Section 4.2.4), as well as that the tight timelines reduced the scope for adapting the approaches beyond simplifying implementation (see sections 4.1.4 and 4.2.4). That said, there were instances cited where remedial action was taken in response to monitoring data highlighting challenges, although features of this are down to features of the PbR modality (external verification of results) and programme design elements (having a dedicated outcome phase and enhanced accountability for outcomes).

The team found evidence of remedial action being taken in response to challenges on the ground but this was in the context of a limited degree of flexibility at the implementation level. Overall, the PbR modality is not seen to have greatly enabled the implementation level to better respond to challenges on the ground.

Strength of evidence: suggestive/moderate

During the outcome-phase interviews, the engagement with field-level staff was limited. However, the interview feedback from the country programme level is consistent and articulates a plausible mechanism of impact.

[39] Included in the PbR TOC developed by the evaluation team and used in the contribution analysis as a testable proposition.
[40] During the endline data collection the evaluation team had limited sight of the field-level implementation due to restricted travel in Pakistan and because, for the Bangladesh evaluation, only country-level staff were interviewed.

4.2.6 Learning in practice

WEDC facilitated inception workshops with consortium members and IPs in May and June 2014, with one workshop in Bangladesh and one in Pakistan. Similarly, WEDC hosted two learning events with consortium partners and IPs in February 2015, (one in Bangladesh and one in Pakistan) and in March 2018 a learning event was held in London.

At the end of the output phase, an Advocacy and Learning Plan for the outcome phase was adopted and at the time of the Pakistan case study 13 case studies were under development. SAWRP I presented at each of the WEDC conferences during implementation, and participated in the 'tri-partied[41]' WRP event at World Water Week in Stockholm in 2018.

WEDC had a role in supporting knowledge management and learning. Some SAWRP respondents in Pakistan viewed learning as slightly marginalised as a result of the PBR modality. Similarly, in Bangladesh programme managers at country level acknowledged that not much was done in terms of structured learning over the course of the programme beyond the production of some case studies. Plan International Bangladesh did highlight, however, that experiences were shared at WASH conferences[42]. The reason cited for the general marginalisation of learning was that programme staff were preoccupied with quarterly reporting and meeting contractual targets. Some respondents felt that a better balance of 'hard' and 'soft' programming should be established in future.

Nevertheless, there was a great deal of learning through experience given that country partners were implementing a large-scale WASH PbR programme for the first time. Part of this learning was on programme monitoring, and the online system developed by Plan International Bangladesh was shared internally across the organisation and adopted by the health team. The partnership models adopted under the programme were also cited as an area of learning. In particular, staff from both countries expressed their appreciation for the skills-based training and up-skilling (e.g. mobile monitoring and project management) that occurred throughout the programme.

In discussing its role, WEDC highlighted that over the course of the WASH Results Programme the focus of its role shifted. Earlier in the programme the focus was on supporting developing the monitoring frameworks, while towards the end the focus was more on learning. While this sort of evolution could be expected there was a sense that earlier in the programme support to monitoring 'crowded out' the learning function as the monitoring and evaluation requirements were greater than expected.

The findings highlight that the majority of the learning under the programme was within the consortia – and that at times the pressure to deliver results marginalised a focus on learning. That said, there are clear examples of where the wider sector was engaged, while at the time of writing there are still learning products under development by SAWRP.

[41] The Suppliers, DFID, and MVE teams
[42] South Asia Conference on Sanitation (SACOSAN) VI in 2016.

Strength of evidence: moderate

Clear feedback from supplier leads and the learning partners as to what the focus of the learning efforts were as well as the key initiatives/products.

4.2.7 Significance of external factors

In both countries there were factors that affected implementation, although not to the level that renegotiation of contracts was required. SAWRP staff highlighted several significant external factors that affected implementation:

- **Environmental shocks** - In both Pakistan and Bangladesh heavy rains and flooding reportedly destroyed latrines in both the output and outcome phases. For example, in the output phase Plan International Pakistan and its IPs reported that 22% of latrines in Umerkot, Sindh (and a few communities in Punjab) were damaged by heavy rains and required reconstruction. No extra funds were available for the reconstruction effort, but it was the responsibility to mobilise the necessary community action to repair the damage: *'We had to reconstruct 4,700 latrines*[43] *destroyed during heavy rains last year, but managed to meet our targets' (LPP SAWRP, Project Manager, and Lodhran).* In Bangladesh it was reported that the SAWRP 'contingency funds' were used as part of the response to flooding.
- **In Bangladesh in 2015 there was a political crisis** following the 2014 election and this was accompanied by widespread protests and strikes by government staff. This was particularly a challenge for Plan International Bangladesh, which was working directly with government staff in implementing the programme. Staff at the country management level also highlighted that frequent rotation of government staff posed a challenge as each new post holder needed orientation.
- **The sterling devaluation related to Brexit** was highlighted as impacting on the budgets in both countries and for some partners it was necessary to draw on the 'contingency fund' that had been built into the SAWRP budget from the outset. This was also a factor in the decision by WaterAid headquarters to reportedly allocate roughly £400,000 of additional funds from its unrestricted resources. Plan International reported that their output-phase activities cost less than expected and this in part enabled them to keep IP staffing levels higher than originally planned during the outcome phase, despite the Brexit-related devaluation of the pound.

Strength of evidence: moderate

We are confident that the key external factors affecting implementation are understood and discussed where appropriate.

4.2.8 Extent of attribution

The methodology for the programme evaluation means that the team were unable to establish a quantitative counterfactual against which to compare progress, although parallel to the programme evaluation a separate impact evaluation was conducted that focused on the effectiveness of the outcome-phase activities in two of the provinces in Pakistan. These workstreams were designed to

[43] The burden on the IP is the time and effort it took to re-mobilise the community and not in terms of material costs.

be separate, and the findings of the impact evaluation are reported separately, but the findings from the impact evaluation do have some implications for the programme evaluation findings. A summary of the key findings from the impact evaluation is given in Box 7.

Box 7: Summary findings of the impact evaluation in Pakistan

The Randomised Control Trial (RCT) Research Study focused on SAWRP in Pakistan, and in particular on three out of eight districts covered by SAWRP in Pakistan, located in the provinces of Punjab and Sindh. It did not seek to address questions related to the PbR modality but rather focused on the *additional* benefit of the outcome-phase activities with respect to key outcome-level indicators. The primary research question of the RCT was:

What is the impact of outcome-phase activities on the sanitation behaviour of output-phase beneficiaries?

A mixed-methods approach, consisting of a quantitative and qualitative component, was chosen to answer the research question. Under the quantitative component the research key research questions was addressed using experimental methods (specifically an RCT). In order to be able to gain a deeper understanding and to explain patterns, trends, and mechanisms (the 'what' and the 'why'), which cannot be readily addressed with an RCT, the quantitative component was coupled with a complementary, sequential qualitative research study.

With regards to the primary research question the RCT Research Study found that, on average, the outcome-phase activities did not reduce this slippage at a statistically significant level, although in Sindh the programme activities successfully reduced slippage by 14 percentage points. This impact seems to have been achieved by reducing the degree to which toilets became dysfunctional. Latrine collapse due to rains was reported as a key reason for latrines becoming non-functional in communities in Sindh, but less so in treatment communities. In Punjab there was limited slippage across treatment and control, and therefore limited room for the outcome-phase activities to have an impact.

Beyond the findings of the impact evaluation this section considers where other resources were explicitly leveraged by SAWRP towards the achievement of outcomes, other factors that may have contributed to the programmes results, and the significance of the monitoring and verification approach to considering attribution.

In Pakistan the programme was mostly implemented by SAWRP partners, with few instances of other organisations contributing directly to the outputs and outcomes. This said, the programme worked through a large network of community-level volunteers and local government was involved in some promotion activities and ODF certification. These people's time is not reflected in programme expenditure or the prices per beneficiary figures agreed with DFID. In Bangladesh, Plan International worked directly with local government staff in implementation but their inputs were not factored into SAWRP prices.

Both the lead and implementing partners reported that at times they drew on other organisational funds to support implementation, but the evaluation team has not had sight of the details.

Strength of evidence: suggestive

The lack of detailed financial data and a quantitative counterfactual hampers this assessment. The RCT findings in Pakistan provide robust evidence for assessing the relationship between the outcome-phase activities and results through a quantitative counterfactual, but these results are rooted in their context and cannot be generalised across the *two countries in the programme.*

4.3 Efficiency

Box 8: Overall evaluation questions related to this section's discussion

DEQ 3.1: How efficient was the tendering and procurement process and what effect did this have on programme delivery?

DEQ 3.2: To what extent were the individual sub-programmes designed and delivered in a cost-efficient and cost-effective manner?

DEQ 3.3: Under which circumstances did the PbR modality affect the cost-efficiency and cost-effectiveness of individual sub-programmes?

DEQ 3.4: Under which circumstances did the PbR modality strengthen the programme monitoring and management arrangements of individual sub-programmes?

DEQ 3.5: Under which circumstances did key programme features affect cost-efficiency and cost-effectiveness?

DEQ 3.6: Under what circumstances did consortium complexity affect the efficiency of the programme management arrangements of individual sub-programmes?

DEQ 3.7: To the extent that new PbR risk-sharing arrangements were applied within consortia, how did this affect programme delivery?

4.3.1 Approach to risk sharing and programme implementation

Between the three consortium leads there was a PbR agreement relating to output-level results whereby partners were paid *pro rata* against results. Within SAWRP, most of the consortium partners worked with local IPs.

In all cases these partners were on grant agreements. In Pakistan they were managed tightly and SAWRP planned for a level of overachievement as a risk-management strategy. Similarly, in Bangladesh – where only WaterAid worked with local partners – the partners were also on grant agreements. As in Pakistan they were managed tightly; however, unlike in Pakistan WaterAid introduced a system of penalties and upside incentives to reward overachievement with a contracting modality that had elements of PbR. If the IPs underachieved against target (~<80%) there was a financial deduction, although there were upfront payments to local partners. There were also upside incentives: partners received full payment at 90% of the target being achieved, and then beyond that received 'points' for levels of overachievement (where one point was equivalent to roughly £1,000). WaterAid Bangladesh planned for 5% overachievement in all grants as a risk-management strategy. Country management reported that the penalty was used in some occasions – albeit rarely – and there were lots of upside payments due to overachievement. Apart from the challenges that pre-financing would create for local NGOs, one respondent highlighted a hesitance to adopt a PbR modality with local partners as it was seen to create incentives for the IPs to cut corners, although the same argument could of course be applied at supplier level.

Another important dimension to the risk sharing within the consortium was that the financial risk was held at the headquarters' (UK) level. As such the country offices were insulated from the risk of large financial penalties, although they nonetheless felt great responsibility for meeting the targets. While the PbR modality presented financial risks for the CPs, these proved to be manageable and, importantly, were managed at global rather than country level. Plan International and WaterAid headquarters were able to maintain a regular flow of funding, while both country programmes had other, grant-funded programmes in their portfolio that helped to even out peaks and troughs in the funding of overheads. Plan International, for example, had a DFAT-funded sanitation programme running concurrently with SAWRP, also on a fairly large scale. This meant that the suppliers could, on occasion, draw on alternative funding sources to plug gaps. This did not eliminate risks entirely, however; during the midline case study one programme manager remarked that *'We don't have that kind of funds available to deal with huge financial backlogs and it certainly creates pressure on us'* (SAWRP Project Manager, Plan).

During the outcome phase the partners' results were jointly assessed, in effect pooling the risk and responsibility for the results. This was significant as in the Q1 2017 payment round there was a 14% payment deduction for underachievement on hygiene outcomes. Arrangements for apportioning such deductions were not defined in advance. The deduction was applied by the partners in proportion to their share of the results. The 'weighted achievement' against targets was at first found to be 58% across the three component indicators, but this was later revised to an 86% 'weighted achievement' following negotiations with DFID. The justification related primarily to high baseline results and questions surrounding the measurement validity and 'non-sample error'[44] in the baseline data, coupled with the fact that survey respondents were drawn randomly from the community rather than being specific programme beneficiaries. Part of the debate around these results was related to Ipsos MORI's capacity to implement WASH surveys, particularly the design elements related to questionnaires and some of the nuance related to how questions were asked by enumerators.

Strength of evidence: moderate

Contracting arrangements within the consortia, and the key implications for risk sharing, are relatively clear. However, the team lacked access to the specific agreements.

4.3.1.1 Effect of the PbR modality on supplier staff

At the midline assessment in Pakistan all organisations (suppliers and IPs) agreed that they would increase the size of their teams, especially the monitoring component. It was evident from the field visits and interviews with district-level staff that the burden of meeting tight deadlines for submitting monitoring reports, providing evidence, and dealing with other quality assurance issues fell on the shoulders of the monitoring officers. Monitoring staff highlighted working long hours (10–12-hour days), as one of the coping mechanisms used.

During the endline case study in Pakistan the pressure that was placed on staff during the output phase was again raised. For example, LPP highlighted that they changed their office opening hours to allow working into the evenings. Additionally, some IP staff reflected that the long hours

[44] Non-sample error is an umbrella term for measurement in error not attributable to the sample design. Non-sample error cannot be quantified. Most often it relates to communicating concepts and definitions to enumerators and respondents, questionnaire design, implementation approach, and data tabulation and analysis.

and increased pressure was more of an issue for female staff, who had to balance their work and family responsibilities.

Staff in Pakistan at all levels and the country management staff in Bangladesh highlighted that the features of the PbR modality enhanced accountability and focused implementation around results over conducing activities. That said, some argued the lead suppliers already had a strong motivation to deliver results; their priorities were well aligned with DFID's and hence there was no need to introduce additional incentives to do the right thing. There was, in any case, no upside incentive for over-delivery, only the threat of penalties if targets were not met. A related point is that reputational risk was itself a strong motivator for the suppliers beyond the financial incentives.

Strength of evidence: moderate

The link between the PbR incentives and staff attitudes/behaviours was well understood in Pakistan, with clear and consistent messaging across the implementation levels and many staff explicitly highlighting they felt a shift in attitudes. The findings are corroborated by the country manager interviews in Bangladesh, albeit with the caveat that field-level implementation staff were not interviewed in Bangladesh.

4.3.1.2 Effect of the PbR modality on relationships with partners and government

The PbR modality appeared to have no negative impact on relationships with partners. Relationships with IPs were reported as generally cordial, though several respondents in Pakistan[45] expressed frustration with output-phase timelines and the pressure these created on staff.

The local government departments in the three provinces in Pakistan were interviewed. All were familiar with the programme and complementary about the level of coordination. Additionally, some remarked that the credibility of the data due to verification was appreciated as it gave them greater confidence in the results. Nevertheless, PbR was not a significant aspect of the programme for them, with just one interviewee remarking that they saw it as introducing greater accountability and that provincial government in Punjab would be willing to take up such a contract.

These findings are consistent with the midline assessment, which did interview local staff in the field in Pakistan and noted: 'a*t the District level, IPs engaged in their normal way with government structures, in particular, PHED and LGD. The capacity of government structures depends largely on the motivation and capacities of individuals. PHED staff were found to be more active in Bahawalpur District than Lodhran District. PHED staff in Bahawalpur were supported by the UNICEF SPSP programme, while the staff in Lodhran were not. The role of government has been minimal in the output phase, but the role of government becomes more critical in the outcome phase'.*

While institutional strengthening featured in the SAWRP programme design, in practice it did not gain a high profile, partly because results were defined only in terms of beneficiary numbers. In

[45] Note that IPs were not interviewed in Bangladesh.

practice, country partners' engagement with government under SAWRP in Pakistan was mostly at the operational level, i.e. at UC and district level.

In Pakistan the PbR modality did not greatly affect the relationship between suppliers and government staff. The evaluation team is unable to comment on how the PbR modality affected relationships in Bangladesh.

Strength of evidence: suggestive/moderate

The relationship with government is relatively well understood in Pakistan, although the programme was largely implemented independently of government. In Bangladesh, where the programme was implemented in close partnership with government, the relationship is less well understood as the government staff in Bangladesh were not interviewed as part of the evaluation.

4.3.1.3 Relations with DFID

The relationship with DFID was managed centrally by Plan International with input from the country CCUs on reporting.

4.3.2 Efficiency of management arrangements

Working in a consortium imposed its own challenges as it meant that the programme had more administrative layers and reporting lines than would normally be the case. Having a dedicated CCU was reportedly helpful insofar as it took some of the administrative burden off the shoulders of the CPs' permanent WASH managers. At the same time, however, the presence of the CCU was a complicating factor in that CPs lacked a common understanding of its role. While the CCU appeared on paper to be the institutional apex of the country programme, calling it a 'Coordination Unit' rather than a 'Programme Management Unit' signalled that its role was largely administrative, pulling together reports and acting as the communication link between Pakistan and the UK. At times some CP personnel in Pakistan felt that the CCU was over-reaching by asking for too much information on programme activities, while its physical location[46] gave the misleading impression that it had a closer allegiance to Plan International than to WaterAid. In Bangladesh, the CCU had its own office separate from the country partners, but in the outcome phase it moved into Plan International's country office to reduce costs. A number of respondents commented that it would have been better to base the CCU in a neutral location outside of Plan International or WaterAid.

Another challenge to consortium working was that, in Pakistan, the School of Five component led by Unilever was managed from Karachi rather than Islamabad (where the other suppliers were based) and operated for a shorter period than the rest of the programme. Potential synergies at field level were therefore not fully exploited in the limited number of cases where Unilever and Plan International or WaterAid were working in the same sub-districts[47].

As outlined in Section 3.4.3, both Plan International and WaterAid in Pakistan had an existing working relationship from work in flood relief: this, combined with the application of a tested operational model, was cited as a source of efficiency. The need for the CCU was

[46] The CCU was moved to be housed within Plan International as a cost-saving measure in the outcome phase in Pakistan.
[47] The implementation areas intentionally did not overlap to avoid the issue of 'double counting' beneficiary numbers.

questioned by some SAWRP respondents; while others highlighted that, while its role shifted over the course of the programme, this was from necessity and it played a key role in enabling the verification.

Strength of evidence: suggestive/moderate

Clear interview information is available on key areas, but a lack of any financial information constrains the analysis.

4.3.3 Efficiency of programme monitoring

The monitoring requirements were greater than the partners and IPs had anticipated. At the start of the programme in Pakistan both the suppliers and their respective IPs lacked standardised software and forms for data collection. The first two quarters were demanding for the country teams in terms of data collection, particularly as decisions were taken at global level, by the verifier, to ask for additional information such as number of children under five in a household or CNIC[48] (identity card) number. This required IP teams to re-visit villages and households to collect the missing data. With time, the data collection process improved, with IP staff increasingly feeling more confident in dealing with the demands.

While the challenges with the overall data collection frameworks were consortium-wide there was an important difference between the programmes in Pakistan and Bangladesh. The data collection process at the field level in Pakistan was largely paper-based across the programme, although WaterAid later used the mWater mobile-to-web (M2W) platform. During the output phase the results were aggregated into Excel spreadsheets by IPs and the partners. In Pakistan, the monitoring requirements gave rise to the phrase ‘payment by paperwork’.

The limits of Excel as a tool for monitoring a programme of this size were highlighted, particularly as much of the CCU’s time was lost removing ‘cell errors’ from the ‘latrine database’. With hindsight, more time should have been dedicated to designing the data collection and management system in the start of the programme (and a clear inception phase included). . In contrast, in Bangladesh the partners used M2W monitoring software from the beginning. The rollout of this was not exclusively related to this contract though – for example, WaterAid had been working for many years on the development of the mWater monitoring platform[49].

As outlined in Section 3.4.6 the monitoring requirements associated with the outcome-phase payments formed only a small part of the monitoring undertaken by the partners. In Pakistan the monitoring arrangements were such that during the outcome phase most households were likely to be visited at least three times for someone to check the status of their latrine and could be visited as many as six times. Furthermore, the visits would be from different people the community members were unfamiliar with. In the outcome phase, the following visits were possible:

- CRPs as part of the programme monitoring (household highly likely to be visited multiple times);
- social organiser monitoring (household highly likely to be visited during the outcome phase);

[48] Which SAWRP global M&E staff report can be controversial in Pakistan
[49] An open source software platform for monitoring WASH: www.mwater.co/

- IP programme management (visits conducted to a sample of households for quality assurance);
- CP district programme management (visits conducted to a sample of households for quality assurance);
- CP country programme management (visits conducted to a small sample of households for quality assurance);
- CCU monitoring visits (visits conducted to a small number of communities); and
- Outcome phase survey (social organisers from different UCs visit if the household was sampled in one or both of the outcome-phase surveys. Spot /back checks also possible under the survey).

These monitoring visits were not required by the MVOC process but were introduced by the partners as part of their monitoring and programme management process, and as a key risk mitigation strategy against unexpectedly poor survey results. The multiple management levels in part necessitated so many visits and, with the benefit of hindsight, some of the country-level staff reflected that this could have been streamlined. Similarly, we heard multiple stories (from implementing staff) of households expressing frustration with the number of monitoring visits.

WaterAid in Bangladesh adapted some of the monitoring improvements introduced under SAWRP for use in other programmes, including: mechanisms for feeding back the findings of monitoring activities to IPs; using mobile phone-based reporting; making spot checks; quality assuring monitoring systems; and strengthening accountability for the delivery of results. Having said this, one respondent also noted that the programme had generated a mass of outcome data and could have made better use of it to improve programme effectiveness.

The use of similar implementation approaches and a single monitoring framework was also seen as offering economies of scale, though this created its own challenges as the monitoring systems of each of the suppliers across the countries had to be adapted to the common frameworks. In the process of monitoring there were some potential inefficiencies related to the systems and processes used, particularly that the use of M2W systems offers efficiency gains against a combination of paper-based data collection later keyed into Excel databases.

Strength of evidence: moderate

Analysis draws on both the structural elements of the monitoring framework as well as clear information from interviews. However, a lack of any financial information constrains the analysis.

4.3.3.1 Relations with verifiers

Relations with the verifier were reportedly cordial; on the whole the process ran smoothly and feedback at debriefing meetings led to remedial action being taken where necessary. Good communication with IPs also helped to ensure that any issues raised were resolved efficiently. Having said this, at times there were multiple requests for clarification or additional information from e-Pact in the UK and responding to them could be challenging due to the time it took to collect the information.

Verification also involved validating suppliers' work, which many supplier staff remarked was satisfying. In addition, some said that it increased government ownership of the programme. However, in Pakistan[50] the country verifiers did express some dissatisfaction with their role, particularly that there were a limited number of days and that professionally they felt they had more to offer the programme beyond just verifying the results. In future, to make the role more meaningful, it was suggested more field visits should be included as an integral part of the work, as understanding the context is crucial.

The lead verifier and country verifier interviewed in Pakistan also felt frustrated by not being able to provide technical advice to the 'supplier' given their own wealth of WASH experience. That said, it is arguable that if the verifiers had a greater role in programme implementation this would present a conflict of interest with regards to the monitoring function.

Strength of evidence: moderate

Clear information was available from both the suppliers and the verifiers.

4.4 Impacts

Box 9: Overall evaluation questions related to this section's discussion

DEQ 4.1: How likely is it that the programme will achieve its health and non-health impacts?

DEQ 4.2: Under which circumstances did the WASH Results Programme activities have any unintended/unplanned positive or negative impacts?

4.4.1 Prospects for health impacts[51]

The evaluation design recognises that it was not possible to measure the actual health impacts of the WASH Results Programme. As a result of this lack of data, the evaluation assessed the *prospects* for health impacts by addressing four questions:

To what extent have services and behaviours continued to function and be used since their initial implementation (sustainable outcomes)?

The outcome-phase survey results (see Section 3.3) across both countries were positive, with both overachievement against the target as well as high overall functionality (especially with regards to water and sanitation).

To what extent have the utilisation of water and sanitation services and the uptake of hygiene practices reached all members of target populations (inclusive outcomes)?

The achievement of ODF was not a result linked to payment. Nonetheless, the partners in Pakistan highlighted that it was a key aim in their programme approach and reported high levels of ODF certification by government to the evaluation team. However, it should be noted that in Pakistan

[50] Note, again, that the country verifiers in Bangladesh were not interviewed.
[51] The discussion in this section pertains only to Pakistan as there was insufficient information to comment on Bangladesh.

the criteria and process for certifying ODF varies across and within districts, meaning the indicator and reports from partners are difficult to interpret. The outcome data reported by SAWRP was exclusively drawn from a sample of the output-phase beneficiaries and does not provide a view into the community-wide sanitation status. Ultimately, there are inadequate data to comment on the extent to which benefits were community wide. However, as will be discussed in the section below, the team note that there were positive design features that supported an inclusion focus.

To what extent has the programme advocated for, and successfully influenced, attempts to achieve sustainable WASH services across entire districts (or beyond)?

If the use and maintenance of latrines reaches across entire communities and is sustained then this will contribute to safeguarding the health of the local population, as will the continued consumption of adequate, safe water. SAWRP results seem to be somewhere in the middle in terms of their geographical reach; many of the ODF communities are village clusters rather than entire revenue villages and so are adjacent to, perhaps surrounded by, communities which are not ODF and therefore pose a continuing risk to health. However, there are also some programme districts where SAWRP has been complemented by other WASH projects and the majority of rural communities in the district are now ODF. One example is Bahawalpur District in Panjab, where UNICEF's partners also implemented PATS at scale from 2014 onwards, under DFID's ASWA programme.

What other obstacles exist to the realisation of the full potential health benefits of the WASH programme, in areas such as nutrition, shelter, livelihoods, and education?

With the data available to the evaluation team it is not possible to assess the extent of potential other obstacles to health impacts.

Strength of evidence: suggestive

Neither the programme nor the evaluation sought to measure impact-level indicators. The experimental evidence relating to the relationship between WASH outcomes and health impacts is developing and there is an increasing body of evidence supporting which factors are most significant. Nonetheless, the evidence base remains incomplete and there are significant areas of uncertainty.

4.4.2 Equity and inclusion

The framework for assessing how effectively the programme addressed equity and inclusion was developed during the design phase of the evaluation and refined ahead of the endline case studies. The logic and justification for the indicators included is contained within an annex to the 'Evaluation Design Document'. The framework was only applied in Pakistan, as the remote interviews with the Bangladesh programme managers did not provide enough insight to reliably complete the framework.

4.4.2.1 Extent of focus on inclusion

Table 20 presents a summary 'equity prerequisites' matrix for the Pakistan programme overall, while Annex A contains further details. Generally speaking, the programme scores well in terms of

its inclusive and pro-poor approach, although we are unaware of any specific interventions on this at policy level (provincial or national).

Table 20: Framework used to assess equity focus – Pakistan

Areas of investigation	Achievement
Programme planning and implementation	
1. Within targeted locations, did the programme endeavour to meet the needs of all, including communities that were harder to reach or serve?	Yes
2. Was technology selection (where relevant) and detailed design undertaken with the full participation of the intended beneficiaries?	Somewhat
3. Within targeted communities, did operational approaches address the needs of marginalised groups/households and of those with physical disabilities and infirmities?	Somewhat
4. Did women participate actively in programme implementation and were they adequately represented in decision-making processes?	Somewhat
Monitoring	
5. Did monitoring at output level generate disaggregated beneficiary data confirming that the programme provided access to WASH facilities for marginalised groups and those with special needs?	Somewhat
6. Did outcome-phase surveys confirm the use of WASH facilities and adoption of hygienic behaviour by marginalised groups and those with special needs?	Somewhat
Addressing institutional barriers	
7. Where relevant, did the suppliers, in collaboration with other development agencies, work to strengthen the policy and institutional environment for equity and inclusion?	Unlikely
8. Where discriminatory practices existed within government institutions, did the suppliers advocate for a more inclusive approach?	n.a.

Partners gathered some data on disability as part of their routine monitoring and during the outcome surveys, but this was not reported to DFID. Two dedicated assessments two dedicated on equity (midline and endline) were conducted by independent consultants under the SAF. At the time of the evaluation visit, the results of the second round of SAF surveys were not yet available. However, the first equity survey found that the programme had generally performed quite well in this area. For example, the programme featured measures to:

- involve marginalised groups in planning and decision making;
- ensure that communal facilities met the needs of people with disabilities; and
- involve children in sanitation and hygiene promotion and encourage them to be change agents.

The report also noted some challenges, however. For example, Plan International needed to provide training on equity and inclusion for programme staff (which they subsequently did) and there was weak understanding of equity and inclusion among government partners.

Even without access to the second study, it is evident that SAWRP has targeted some of the poorest and most under-served communities in the provinces where it operates. WaterAid Pakistan, for example, had made a decision prior to SAWRP to focus their support on Rajanpur

District in Panjab, plus Badin and Thatta in Sindh, on the basis that they had some of the lowest levels of WASH access in those provinces. WaterAid Pakistan reportedly targeted these districts even though they were susceptible to periodic flooding, a factor which apparently deterred some other development agencies.

Programme strategy also featured measures, based on the PATS guidelines, to enable the poorest and most disadvantaged within each targeted community to access improved sanitation facilities. The operational approaches adopted by both Plan International and WaterAid included the establishment of sanitary marts offering easy access to affordable materials and skills for latrine construction, while WaterAid also provided free 'demonstration' latrines to a small number of very poor households in each community (though Plan chose not to do this).

Although the equity analysis framework was not applied to Bangladesh, the SAF studies there (which were conducted by an independent consultant)[52] provided some useful summary findings:

> *'The principles of equity and inclusion are firmly embedded within the design of the SAWRP programme, and in the implementation approaches used, especially when working with communities. However, these approaches and processes are not as well documented by Plan International and WaterAid in Bangladesh or their partners, compared to other programmes they have managed. This occurred due to the PbR approach adopted on this project, which focused on achieving large numbers of beneficiaries within a tight timeline in the output phase, which made it difficult to reflect on equity findings and then improve processes going forward. This approach has a tendency to lead to a "one-size fits all" approach to implementation. Better documentation of processes and approaches would support stronger programmatic learning and future implementation. Dissemination of that key learning also needs to be planned and implemented'.*

During interviews with the country partners in Bangladesh it was reported that they sought to ensure that the needs of people with disabilities were met, despite there being no special provision for this within the programme budget. For example, WaterAid Bangladesh hosted a meeting with people with disabilities to better understand their WASH-related needs and offered a range of relevant latrine options. Furthermore, both partners community engagement approach included a community-led process of identifying both people with disabilities and the 'hard-core' poor who warranted support with latrine construction or improvement. Country partners also sought to ensure that the poorest community members participated in the meetings, and thus held some of them in the evening when people were not working. The implementation process also sought to ensure that women and girls were adequately represented on village committees and school health clubs, and in decision making generally, so that WASH facilities installed or improved under the programme responded to their needs and priorities.

A similar approach was adopted by WaterAid in Pakistan. During the output phase, the needs of PWD were identified during community meetings linked to the triggering process, and in the outcome phase, quarterly broad-based community meetings provided another opportunity to check that the needs of PWD were being addressed.

Overall, the evaluation notes positive features in the programme's design and implementation that are likely to have supported inclusion, though there was limited

[52] We have not seen a report on the second-round SAF studies, if any exists.

monitoring and reporting related to DFID on inclusion. Within the programme design much of the inclusion focus is assumed; more explicit reporting in this area is important to corroborate these assumptions.

Strength of evidence: suggestive/moderate

There is limited quantitative evidence from either the programme or the evaluation that accurately describes the intra-community or intra-household distribution of benefits. However, there is reasonable evidence supporting the elements of the assessment related to targeting and programme design elements.

4.4.2.2 Effect of PbR on inclusion focus

The PbR modality is seen to have not significantly affected the programme's strategy on inclusion. The focus on inclusion within the programme was largely rooted in the organisational values, with the partners undertaking many initiatives to strengthen inclusion outside of the results tied to payment. This is perhaps most explicit in the case of ODF status in rural sanitation, which by definition focuses on community-wide outcomes. Similarly, the structure of the payments and the indicators used were not seen to have explicitly incentivised an equity focus.

Strength of evidence: moderate

Given the caveats related to quantitatively assessing the level of equity, there was consistent interview evidence on how the PbR modality affected the equity focus.

4.4.3 Unintended positive or negative impacts

As discussed in Section 4.3.3, staff in Pakistan emphasised that the intensive monitoring sometimes damaged the relationships they had with the community. It was reportedly not uncommon for a household to be visited four or five times for someone from SAWRP to look at, or verify, their latrine during the outcome phase. Part of the confusion or frustration at community level stemmed from the fact that households had not (in most cases) received any material support with latrine construction from the programme, so they did not understand why the NGO had so much interest in their facilities. In some cases, the community were suspicious that the NGOs were collecting money for these latrines as part of the survey data collection.

Strength of evidence: suggestive

The absence of significant work at the community level by the evaluation team means this assessment is primarily based on interview evidence, which is ill suited to capturing this information.

4.5 Sustainability

Box 10: Overall evaluation questions related to this section's discussion

DEQ 5.1: To what extent were the individual sub-programmes designed and implemented to maximise the likelihood of achieving long-term sustainable WASH outcomes and impacts?

DEQ 5.2: Under which circumstances has the PbR modality affected the likelihood of the long-term sustainability of the outcomes and impacts?

DEQ 5.3: Under which circumstances have other programme features affected the likelihood of the long-term sustainability of the outcomes and impacts?

DEQ 5.4: Under which circumstances did the WASH Results Programme contribute to enhanced sector learning to inform better evidence-based WASH policy and programming?

4.5.1 Prospects for sustainability

A risk-based framework was developed by the evaluation team to assess the risks to sustainability. As with the equity framework, the logic and justification for the indicators included is contained within an annex to the 'Evaluation Design Document'. The framework was only applied in Pakistan, as the remote interviews with the Bangladesh programme managers did not provide enough insight to reliably complete the assessment. Annex A contains the detailed frameworks, while Table 21 below presents the summary findings. Risks scoring 1–2 are considered negligible risks and colour coded Green; risks scoring 3–4 are considered moderate risks and colour coded Amber; risks scoring 6+ are considered high-risk areas and colour coded Red. The strength of evidence supporting the assessment is included in parenthesis and is High (H), Medium (M), or Low (L).

The assessment highlights that, while much has been done to consolidate latrine use during the outcome phase, as well as to promote simple, affordable technologies that can be maintained at community level, the enabling environment for sustainability remains weak in terms of ongoing support and monitoring by local government agencies beyond the outcome phase. Particular issues here are inadequate operational funding at local level and a lack of clarity over the roles and responsibilities of specific government agencies[53].

Accountability for outcome-phase results incentivised suppliers to ensure the continued functionality and use of water and sanitation facilities up to the end of the outcome phase. However, this is not attributable exclusively to the PbR modality; it seems highly likely that CPs would have done the same under grant funding if they were held accountable for achievements at outcome level.

CPs recognised the importance of developing the capacity and motivation of government to provide a long-term, supportive role to communities once SAWRP had ended. However, the PbR framework defined results only in terms of beneficiary numbers and this tended to marginalise programme efforts on institutional strengthening.

[53] Notably the overlapping mandates of the Local Government Department and Public Health and Engineering Department.

At least two of the IPs were expecting to continue operating in some of the programme locations after SAWRP had ended. There was potential scope, therefore, for the NGOs to provide further motivational and supportive inputs to some communities (for example, by encouraging maintenance or repairs when latrines filled or were damaged by floods). However, much will depend on the motivation of these organisations to do so, the funding available, and the extent to which they can accommodate this additional work alongside their other commitments. Moreover, the programme sought to cultivate a strong sense of ownership within beneficiary communities so that they would not expect further assistance from the NGOs in future.

Table 21: Summary of sustainability risks – rural sanitation and hygiene in Pakistan

Aspect	Areas of investigation	Risk[54] (1–9)
User / community level		
Functional	1. Are the selected technologies and systems fit-for-purpose and fit-for-context?	2 (H)
	2. Is the construction quality of physical infrastructure adequate?	4 (L)
Institutional	3. Are the responsibilities of service users and support organisations clearly and appropriately established?	2 (H)
	4. Are service users organised, trained, and equipped to undertake management tasks of which they are competent and capable?	2 (M)
	5. Do service users have the means and mechanisms to report faults and request technical assistance?	N/A
Behavioural	6. Has the programme achieved its outcome-level targets? (latrine use; adoption of handwashing with soap; and (where relevant) consumption of safe water)	3 (H)
	7. Has there been substantive action during the outcome phase to consolidate latrine use and the adoption of handwashing with soap?	3 (H)
Financial	8. Did service users make a substantial capital cost contribution? (for household sanitation, this should be the full capital cost barring cases of exceptional hardship)	2 (H)
	9. Is there real demand for the services developed, demonstrated through use and payment of operating / repair / replacement costs?	3 (H)
	10. Will funds collected meet the full lifecycle costs? If not, are arrangements in place for the shortfall to be met by local government or another permanent organisation?	N/A
Environmental	11. Has the long-term adequacy of the quality and quantity of water resources been assessed and, if necessary, addressed? (including the possible impact of sanitation)	N/A
	12. Have the potential impacts of climate change been assessed and addressed in technology choice and system design?	N/A
Equity	13. Have the prerequisites for achieving inclusive WASH outcomes been addressed by suppliers?	2 (H)
Local government level		
Institutional	14. Is external support and guidance (from local government and/or the private sector) accessible and responsive to service users' needs?	6 (H)
	15. In the case of emergencies (e.g. floods), does local government have response arrangements in place to restore services as promptly as possible?	9 (H)

[54] Risk is calculated as the product of the likelihood and consequence scores.

Aspect	Areas of investigation	Risk[54] (1–9)
	16. Do local governments maintain accurate registers of physical assets within their administrative areas, and are asset management plans in place?	3 (H)
Financial	17. Are goods (e.g. spare parts, sanitary hardware) and support services affordable to service users?	4 (H)
National level		
Institutional	18. Are sustainability commitments and plan actions incorporated into sector strategy?	4 (H)
	19. Is there clarity on the monitoring, management, and financing responsibilities of service users, government (each tier), NGOs, donors, and the private sector?	6 (H)
	20. Are sufficient funds transferred from national to local government to enable community support and the active monitoring of WASH services?	6 (H)
	21. Where necessary, are adequate measures in place to develop the capacity of government agencies to play an effective role in service delivery or community support?	3 (H)
	22. Is a viable sector monitoring system in place or under development?	6 (H)
	23. Are measures in place to facilitate learning on sustainability, and the application of that learning?	4 (H)

Box 11: SAF findings in Bangladesh

While the sustainability risks framework was not completed for Bangladesh there are some comments that can be made based on the review of sustainability under the SAFs, which were based on independent studies commissioned by the consortium. This combined with other sources builds a picture across key areas:

- ***Functionality –*** Generally high levels of functionality of project sanitation and water facilities are evident in the outcome survey results. However, non-functionality is likely to increase over time, so efforts are needed to sustain functionality beyond the programme. This includes working with other stakeholders who will support sustaining outcomes beyond the project period. Plan International consider that water point functionality under SAWRP has been similar to what was achieved under their earlier WASH programmes, though better than the national average.
- ***Environmental –*** The geographic footprint and volume of facilities constructed under SAWRP is small scale and therefore any adverse impacts are expected to be limited. Final disposal of sludge from toilets will be the main environmental concern. Rural pit emptying services are rare and very expensive when available. In some remote areas, composting of waste for agricultural use is being practised, but safe handling practices need to be significantly improved through education and behaviour change.
- ***Institutional and financial aspects*** – The findings of the SAFs in this area were positive. The SAF consultant recognised Plan International and WaterAid's support to strengthening existing structures but highlighted that they could invest more effort in re-energising or creating new structures (including user groups, children's groups, WATSAN committees, etc.). Additionally, the high turnover of local government officials due to elections meant that re-training and re-engagement was often necessary.

With regards to rural sanitation and hygiene in Pakistan, the assessment highlights that the majority of the risks to sustainability relate to the institutional capacity and motivation of government to support functionality and use. While the partners made clear efforts to strengthen the capacity of local government it remains debatable as to the level of impact a

programme of this character (centrally managed, results focused, and using an NGO delivery channel) could reasonably be expected to have on institutional capacity.

Strength of evidence: moderate

The picture is clear for Pakistan and based on both the evidence of the outcome survey results and interviews. However, the evidence for Bangladesh is more limited.

4.5.1.1 Effect of PbR on sustainability

The outcome phase strongly incentivised partners to maintain functionality in programme areas until the end of the programme, and it is the view of the evaluation team that it is likely that functionality would have been lower had there not been the outcome phase and payments related to outcomes.

Many implementation-level staff in Pakistan remarked that the outcome phase was a departure from normal programming and led them to focus more on activities that supported instilling long-term behaviour change over focusing only on outputs. However, the team note that most of the activities remained similar in character to those conducted as a matter of course in many WASH programmes (e.g. household follow-up visits, capacity building to local government, etc. – see Section 3.4.4 for an overview of the outcome-phase strategy in Pakistan).

Strength of evidence: moderate

There was consistent interview evidence regarding how the PbR modality affected the approach to sustainability.

4.5.1.2 Dissemination and learning

At the time of the evaluation visit, the programme in Pakistan had not (so far) been actively involved in sharing lessons from SAWRP with national government or other sector stakeholders in Pakistan. Nonetheless, the provincial governments were well aware of the programme and there were some jointly published project summary briefs highlighting successes. Plan International, WaterAid, and their partners nevertheless continued to be active supporters of district WASH forums and suppliers had ongoing dialogue with key provincial government agencies on WASH matters, although this was not related to SAWRP specifically.

In Bangladesh, at the time of writing WaterAid had not done much to document and disseminate lessons from programme experience to a wider audience in the sector. Plan International was similar in terms of documentation but had shared its experience in a few global and regional for a, including SACOSAN in 2016, the 2015 WEDC conference, and a national conference on PbR.

It should also be noted that at the time of writing there are public-facing publications in development from the learning partner (WEDC).

Strength of evidence: moderate

There has been clear feedback from supplier leads and the learning partners as to what the focus of the learning efforts were, as well as the key initiatives/products.

-

Annex A Evaluation questions

Table 22: Evaluation questions

Evaluation question	Midline	Endline	Evaluation method
High-level evaluation question 1 (*HEQ1*) *Relevance: Were the programme objectives appropriate, and achievable, and to what extent was DFID's programme design and the consortium sub-programme design consistent with achieving these objectives?*			
DEQ 1.1: To what extent were the programme objectives clearly articulated?	✓		Document review
DEQ 1.2: To what extent does the programme's design (i.e. the TOC) set out a clear and realistic process for how programme activities will achieve the intended outputs, outcomes, and impacts?	✓		Document review
DEQ 1.3: To what extent were the scale and pace of the programme (including the December 2015 deadline) realistic for achieving intended outputs and outcomes given the capacity of suppliers and their local partners?	✓		Document review
DEQ 1.4: To what extent was the PbR modality appropriate for achieving sustainable and inclusive WASH outcomes, given the capacity of suppliers and the timeline of the programme?	✓		Process evaluation
DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?	✓		Critique of TOC
DEQ 1.6: How likely was it that the programme design would encourage suppliers to propose 'innovative WASH interventions'?	✓		Critique of TOC
DEQ 1.7: How likely was it that the programme design would encourage inclusive outputs and outcomes?	✓		Critique of TOC
DEQ 1.8: How appropriate was the WASH Results Programme's design for achieving the programme 'learning objectives'?	✓		Critique of TOC
DEQ 1.9. To what extent was the design of each consortium sub-programme appropriate for achieving DFID's key objectives?	✓		Critique of TOC
HEQ2 Effectiveness: To what extent and under which circumstances did the programme operate as intended and which factors helped/hindered the achievement of output and outcome objectives?			
DEQ 2.1: Did the programme achieve the intended outputs at scale?	✓		Process evaluation
DEQ 2.2: To what extent have the utilisation of water and sanitation services and the uptake of hygiene practices reached all members of target populations (inclusive outcomes)?		✓	Process evaluation
DEQ 2.3: To what extent have services continued to function and have behaviours continued to be used since their initial implementation (sustainable outcomes)?		✓ *Wording updated*	Process evaluation
DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?	✓	✓	Contribution analysis
DEQ 2.5: Under which circumstances did the PbR framework help/hinder the achievement of intended outputs and outcomes?	✓	✓	Contribution analysis
DEQ 2.6: Under which circumstances did the PbR framework affect the quality of programme implementation (positive or negative)?	✓	✓	Contribution analysis
DEQ 2.7: Under which circumstances did suppliers implement innovative approaches and focus on learning?	✓	✓	Process evaluation
HEQ3 Efficiency: Has the programme been designed and implemented in a cost-efficient manner?			
DEQ 3.1: How efficient was the tendering and procurement process and what effect did this have on programme delivery?	✓		Process evaluation

DEQ 3.2: To what extent were the individual sub-programmes designed and delivered in a cost-efficient and cost-effective manner?		Removed	n/a
DEQ 3.3: Under which circumstances did the PbR modality affect the cost-efficiency and cost-effectiveness of individual sub-programmes?	✓	Removed	n/a
DEQ 3.4: Under which circumstances did the PbR modality strengthen the programme monitoring and management arrangements of individual sub-programmes?	✓	✓ *Wording updated*	Process evaluation
DEQ 3.5: Under which circumstances did key programme features affect cost-efficiency and cost-effectiveness?	✓		
DEQ 3.6: Under what circumstances did consortium complexity affect the efficiency of the programme management arrangements of individual sub-programmes?	✓	✓ *Wording updated*	Process evaluation
DEQ 3.7: To the extent new PbR risk-sharing arrangements were applied within consortia, how did this affect programme delivery?		✓ (New EQ)	Process evaluation
HEQ4 Impact: How likely is it that the programme will achieve its final impact objectives while minimising unintended negative consequences?			
DEQ 4.1: How likely is it that the programme will achieve its health and non-health impacts?		✓	Process evaluation
DEQ 4.2: Under which circumstances did the WASH Results Programme activities have any unintended/unplanned positive or negative impacts?		✓	Impact assessment
HEQ5 Sustainability: How likely is it that the WASH outcomes achieved by the programme will be sustained beyond the end of the programme in 2018?			
DEQ 5.1: To what extent were the individual sub-programmes designed and implemented to maximise the likelihood of achieving long-term sustainable WASH outcomes and impacts?	✓	✓	Process evaluation
DEQ 5.2: Under which circumstances has the PbR modality affected the likelihood of the long-term sustainability of the outcomes and impacts?		✓	Process evaluation
DEQ 5.3: Under which circumstances have other programme features affected the likelihood of the long-term sustainability of the outcomes and impacts?		✓	Process evaluation
DEQ 5.4: Under which circumstances did the WASH Results Programme contribute to enhanced sector learning to inform better evidence-based WASH policy and programming?		✓ *Wording updated*	Process evaluation

Annex B (overall) Framework for assessing the prospects for inclusive WASH: SAWRP

Areas of investigation	Achievement			Justification	Strength of evidence
	Low (1)	Medium (2)	High (3)		
Programme planning and implementation					
1. Within targeted locations, did the programme endeavour to meet the needs of all, including communities that were harder to reach or serve?			✓	PATS is designed to eradicate open defecation community-wide and suppliers adopted measures to facilitate affordable access for all to improved facilities	H
2. Was technology selection (where relevant) and detailed design undertaken with the full participation of the intended beneficiaries?		✓		In Pakistan there is a strong preference for pour-flush facilities and these were promoted by the programme. There was not, however, much new design work as popular options were already tried and tested	M
3. Within targeted communities, did operational approaches address the needs of marginalised groups/households and of those with physical disabilities and infirmities?		✓		Affordability issues were given more attention than the needs of people with disabilities	M
4. Did women participate actively in programme implementation and were they adequately represented in decision-making processes?		✓		The evaluation was unable to investigate this directly, but SAF study findings were generally positive. However, they highlighted a need for staff training on equity and inclusion; this was subsequently provided	L
Monitoring					
5. Did monitoring at output level generate disaggregated beneficiary data confirming that the programme provided access to WASH facilities for marginalised groups and those with special needs?		✓		Not part of routine reporting to DFID. Although data related to gender, age, and ethnicity were collected by Plan and WaterAid as part of their social mapping process.	H
6. Did outcome-phase surveys confirm the use of WASH facilities and adoption of hygienic behaviour by marginalised groups and those with special needs?		✓		Findings were positive. Some disaggregation in the Ipsos surveys.	L
Addressing institutional barriers					
7. Where relevant, did the suppliers, in collaboration with other development agencies, work to strengthen the policy and institutional environment for equity and inclusion?	✓			No specific policy-level interventions on equity under SAWRP	H
8. Where discriminatory practices existed within government institutions, did the suppliers advocate for a more inclusive approach?	N/A			Not directly relevant to SAWRP since the focus was on household facilities built and maintained by users at their own expense	M

Annex C Sustainability risk framework: SAWRP – rural sanitation and hygiene

Aspect	Areas of investigation	Likelihood (1–3)	Consequence (1–3)	Risk[55] (1–9)	Justification	Strength of evidence
User / community level						
Functional	1. Are the selected technologies and systems fit-for-purpose and fit-for-context?	1	2	2	Low-cost, tried and tested latrine designs were promoted	
	2. Is the construction quality of physical infrastructure adequate?	2	2	4	The evaluation could not investigate this directly but suppliers report that almost all toilets were built or upgraded to JMP standards. Quality of superstructure not important so long as it can be repaired/replaced easily and cheaply	
Institutional	3. Are the responsibilities of service users and support organisations clearly and appropriately established?	2	1	2	Household responsibility for the maintenance of private facilities is generally accepted. Users would not need much assistance in short to medium term, but in the long term would need access to pit emptying services; these are not widely available in rural areas	
	4. Are service users organised, trained, and equipped to undertake management tasks of which they are competent and capable?	1	2	2	Management tasks minor and managed at household level except for FSM	
	5. Do service users have the means and mechanisms to report faults and request technical assistance?	N/A				
Behavioural	6. Has the programme achieved its outcome level targets? (latrine use; adoption of handwashing with soap; and (where relevant) consumption of safe water)	1	3	3	'Use' targets for water and sanitation achieved and verified; final hygiene results pending but largely achieved at midline	
	7. Has there been substantive action during the outcome phase to consolidate latrine use and the adoption of handwashing with soap?	1	3	3	Much has been done on this	

[55] Risk is calculated as the product of the likelihood and consequence scores.

Financial	8. Did service users make a substantial capital cost contribution? (for household sanitation, this should be the full capital cost barring cases of exceptional hardship)	1	2	2	Vast majority of toilets were built without any external support	
	9. Is there real demand for the services developed, demonstrated through use and payment of operating / repair / replacement costs?	1	3	3	Outcome surveys indicate that latrine use is becoming established as a new norm	
	10. Will funds collected meet the full lifecycle costs? If not, are arrangements in place for the shortfall to be met by local government or another permanent organisation?	N/A			Not relevant – no expectation of generating communal funds for household sanitation	
Environmental	11. Has the long-term adequacy of the quality and quantity of water resources been assessed and, if necessary, addressed? (including the possible impact of sanitation)	N/A			Not relevant to rural household sanitation in this case	
	12. Have the potential impacts of climate change been assessed and addressed in technology choice and system design?	N/A			Not relevant to rural household sanitation in this case	
Equity	13. Have the prerequisites for achieving inclusive WASH outcomes been addressed by suppliers?	1	2	2	See Annex 1	
Local Government level						
Institutional	14. Is external support and guidance (from local government and/or private sector) accessible and responsive to service users' needs?	3	2	6	The promotional and supportive role of local government agencies is not clearly defined or funded	
	15. In the case of emergencies (e.g. floods) does local government have response arrangements in place to restore services as promptly as possible?	3	2	6	These services are not generally available from government and there is heavy reliance on external agencies	
	16. Do local governments maintain accurate registers of physical assets within their administrative areas, and are asset management plans in place?	3	1	3	No asset registers for household toilets, but not absolutely essential	
Financial	17. Are goods (e.g. spare parts, sanitary hardware) and support services affordable to service users?	2	2	4	Sanitary marts have been established to support access to sanitary goods in rural areas, but their longevity post-project is not assured	

National level						
Institutional	18. Are sustainability commitments and actions incorporated into sector strategy?	2	2	4	PATS is the unofficial strategy but no specific commitments on ODF sustainability have been adopted at national or provincial level	
	19. Is there clarity on the monitoring, management and financing responsibilities of service users, government (each tier), NGOs, donors, and the private sector?	3	2	6	Institutional responsibilities for rural WASH are not clearly defined	
	20. Are sufficient funds transferred from national to local government to enable community support and the active monitoring of WASH services?	3	2	6	Rural WASH is under-funded by government in all provinces. Though in Sindh, Punjab there are large, externally-funded programmes (or will be soon)	
	21. Where necessary, are adequate measures in place to develop the capacity of government agencies to play an effective role in service delivery or community support?	3	1	3	IPs have provided some capacity building support for local government agencies during the implementation period but resources for longer-term support are not generally available	
	22. Is a viable sector monitoring system in place or under development?	3	2	6	WASH is a provincial subject and none of the provincial governments has a viable sector monitoring system in place, though ODF certification is recorded (on a one-off basis) at district level	
	23. Are measures in place to facilitate learning on sustainability, and the application of that learning?	2	2	4	Some work has been done on this as a multi-stakeholder initiative in selected provinces and at national level, but more is needed	