Evaluation of the WASH Results Programme

Mid-Term Evaluation Report

February 2017

Acknowledgements

Disclaimer

This assessment is being carried out by e-Pact. The project director for the evaluation is Alex Hurrell, with Jeremy Colin leading the Programme Evaluation Workstream (PEW). The project manager is Lucrezia Tincani (Lucrezia.tincani@opml.co.uk).

The contact point for the client is Guy Howard (G-Howard@dfid.gov.uk). The client reference number for the project is PO6507.

e-Pact	Level 3, Clarendon House	Tel	+44 (0) 1865 207300
	52 Cornmarket St	Fax	+44 (0) 1865 207301
	Oxford, OX1 3HJ	Email	admin@opml.co.uk
Registered in England: 3122495	United Kingdom	Website	www.opml.co.uk

Executive summary

The WASH Results Programme

The UK Department for International Development's (DFID's) Water, Sanitation and Hygiene (WASH) Results Programme is a £75 million four-year programme (May 2014 to March 2018) which aims to bring equitable and sustainable water and sanitation services and hygiene practices to 4.5 million people in 12 countries, and thus to improve health by reducing diarrhoea morbidity and child mortality. The programme is implemented by three Suppliers contracted to DFID under a payment by results (PBR) financing modality.

Two of the Suppliers (the Consortium for Sustainable WASH in Fragile Contexts (SWIFT), and the South Asia WASH Results Programme (SAWRP)) are consortia of international non-governmental organisations (INGOs) and others, while the third (Sustainable Sanitation and Hygiene for All (SSH4A) is implemented by a single INGO. All three Suppliers work, to a greater or lesser extent, with national implementing partners, and national and local governments. The three Suppliers each have very different features and approaches.

In addition to the PBR financing modality (in which payments are triggered by the independently verified delivery of pre-specified 'results'), the WASH Results Programme has a number of other key features, the most important of which are:

- **scale**, with each Supplier contract being worth approximately £25 million; and
- **timing and phasing**, with an Output Phase (ending by the Millennium Development Goals (MDGs) deadline of December 2015) that has delivered access to water and sanitation services and hygiene promotion campaigns; and an Outcome Phase (to March 2018), which aims to ensure the sustainability of the services and behaviour changes achieved.

The mid-term evaluation

The WASH Results Programme evaluation is primarily being **undertaken for learning purposes**, not as an accountability exercise. The objectives of the evaluation are to assess:

(i)   whether the programme successfully achieved its stated objectives;
(ii)   the influence of programme design, including the PBR modality, on this achievement; and
(iii)   lessons for applying PBR in WASH programming going forward.

This mid-term evaluation report is the first evaluation output under the monitoring, verification and evaluation (MVE) assignment; it follows the methodology of the Programme Evaluation Workstream (PEW) agreed with DFID in the form of the Evaluation Design Document, submitted in May 2015. The Design Document set out the comprehensive design of the evaluation: the detailed evaluation questions (DEQs) to be addressed, our overarching evaluation approach, and our specific methodology for the various components of the evaluation. The findings to date of the other workstream, the Randomised Control Trial (RCT) Research Study, which examines factors affecting the sustainability of sanitation behaviour change in Pakistan, are described in a separate RCT Baseline Report (May 2016).

This mid-term evaluation examines progress half way through the four-year WASH Results Programme. It answers the sub-set of evaluation questions which relate to the original design and tendering process of the DFID-funded WASH Results Programme, and to its Output Phase (March

2014–December 2015, later extended to March 2016) The remainder of the evaluation aspects will be covered by the endline evaluation report in 2018, at the end of the programme.

The findings of this report draw on analysis of programme documentation from the earliest days of the design phase to date; ongoing analysis of literature on PBR and WASH practice; interviews with key informants among all stakeholder groups; and, in particular, the findings from three purposively selected case studies undertaken in Pakistan, Kenya and Uganda (one country per Supplier). These countries will be revisited at the end of the programme in 2018.

This report will be followed by an endline evaluation report in 2018, at the end of the programme.

Key findings

Programme design

Overall programme objectives are clear, though the theory of change does not articulate how PBR should contribute to programme results. The global design features distinct Output and Outcome Phases, and a strong focus on output targets.

DFID succeeded in stimulating the formation of new organisational partnerships for delivery at scale. At operational level, however, innovation has not been a strong feature of the programme: Suppliers have mostly used familiar, tested approaches, in order to reduce the risk of not meeting targets and deadlines. However, programme management was flexible and responsive to changes on the ground, with many examples of results-oriented problem-solving. The majority of projects were carried out in collaboration with local government agencies or utilities. However, the pressure of both the PBR modality and the MDGs deadline meant that implementation through local government partners was high-risk, and very few projects selected this option.

DFID expected results at scale, and this has been achieved. However, there was no requirement to target entire districts or sub-districts, or to concentrate WASH interventions in the same locations; or to deliver open defecation-free (ODF) communities. Doing so would have enhanced the prospects for health impact. Nevertheless, Suppliers have themselves chosen to prioritise the achievement of ODF status and – for the most part – combined sanitation and hygiene promotion interventions.

The three Supplier programmes differ in terms of PBR contract design and the definition of 'results' that are eligible for payment. With this being a 100% PBR programme, two Suppliers included payments for inputs and processes (i.e. not just outputs) in their results frameworks to enable some cash flow in the earlier part of the programme. This made an already heavy reporting and verification burden even more onerous. The verification process has varied significantly by Supplier, being tailored to individual results frameworks.

Implementation during the Output Phase

The vast majority of output targets were met, with significant over-achievement in many countries. Suppliers had set conservative output targets, given the short time available for the Output Phase and given the fact that the PBR contract by design did not reward over-achievement but penalised under-achievement.

Operational strategies were heavily influenced by (a) the PBR modality (including the requirement for rigorous independent verification); (b) the strict deadline for delivery of beneficiary number

results by December 2015; and (c) the programme 'shape', based on distinct Output and Outcome phases.

While there were understandable challenges given the relative inexperience of the Suppliers and the monitoring and verification (MV) team with regard to the modality, the PBR modality has clear merits. The systems-based verification approach adopted by this programme worked effectively: it led to the strengthening of programme monitoring systems, though the monitoring and reporting burden associated with PBR was felt by Suppliers to be considerable, and in most cases demanded more staff time than the project had anticipated. While it is unclear whether PBR has affected the price-per-beneficiary paid by DFID, PBR has allowed DFID to obtain greater certainty regarding the results achieved, thus increasing the 'value' of what was purchased. Key concerns remain, though, regarding how equity, sustainability and governance aspects can be better embedded within PBR programmes. While most Suppliers did not pass PBR risks down to local partners, the PBR modality still changed the nature of the relationships with local non-governmental organisation (NGO) and government partners to some extent—in some cases placing considerable strain on relationships.

All Suppliers were committed to delivering sustainable WASH services, but pressure to meet output deadlines meant that some projects paid insufficient attention to the ground work for sustainability during the first part of the programme. SSH4A was the only Supplier to build the prerequisites for sustainability into its payment milestones at output level. The non-prescriptive nature of the programme design gave country teams the flexibility to revise implementation approaches where necessary, and this was found to be useful. However, the pressure to meet targets and the MDGs deadline allowed little time for reflection and lesson learning.

The programme has been working in low-income countries, in administrative areas with low coverage of services, and among communities lacking basic services. To this extent, the programme is poverty-focused. What is less clear, however, is the extent to which Suppliers have brought services to the least advantaged individuals, households, communities and administrative areas of the countries involved. The flexibility to transfer beneficiary numbers across consortium members and countries, and flexibility in the use of funds, were appreciated by Suppliers and helped them to ensure that programme targets were met. At the same time, this gives rise to concerns around equity because Suppliers had the freedom to avoid (or move away from) some of the most difficult-to-serve communities and to focus on locations where it was easier to meet targets with the time and resources available.

Looking forward, Outcome Phase strategies covering implementation over the next two years have been agreed, but detailed MV protocols are still under discussion. For water and sanitation, outcome targets have been set at 70%–100% of the output target levels. For the adoption of hand-washing with soap, however, outcome targets have been set lower, at 10%–15% of the output target levels, based on sector benchmarks.

Recommendations

For the Outcome Phase

1. Suppliers should disaggregate their outcome results by gender and wealth quintile, to allow equity to be monitored, and they should specify which implementation approaches will be used to ensure potentially excluded groups are reached.

2. Suppliers should consolidate efforts to develop the capacity of local governments to maintain services developed during the Output Phase.

3. Suppliers should analyse the recurrent costs of sustaining services and work with service users plus local and national governments to agree on the appropriate division of responsibility for cost-sharing.

4. DFID should clarify the advisory function of the MV team, particularly with regard to advising Suppliers on the monitoring framework for sustainability

5. Suppliers and the MV team should prioritise learning so that the programme helps to strengthen capacity for PBR programming in the WASH sector. Specific guidance notes should be produced based on programme learning on monitoring procedures and verification procedures.

For future large multi-country WASH programmes

6. Explicit sustainability requirements should be made so that Suppliers pay adequate attention to sustainability from the start. Explicit measures to promote and enable sustainability should be included in the programme theory of change. Continuous progress should be made towards both outputs and outcomes, from the programme start. Having two phases labelled 'output' and 'outcome' should be avoided.

7. Specific measures to ensure (and monitor) an inclusive approach to programme implementation should be incorporated into the programme.

8. Provision should be made for addressing acute water supply needs by allowing for at least some support to water supply (as was done for SAWRP), so that if a programme finds itself occasionally working in communities where acute water supply problems constrain hygienic practices, these constraints can be resolved.

For future PBR programmes

9. The rationale for using PBR, and its intended effects, should be explicit in the programme design, including the theory of change.

10. The size of the PBR component of programme funding should be reviewed: 100% PBR subjects Suppliers to a high degree of risk given that they have only partial control over the delivery of outcomes and the bulk of programme costs are fixed. We suggest that the PBR element be limited to a small proportion and used to incentivise added value in areas such as quality, sustainability or equity.

11. Programmes should be more prescriptive about which 'results' PBR payments will be linked to: namely, rewarding work in under-served geographical areas; including payments linked to capacity building and systems strengthening; and including payments linked to ODF achievement where feasible. Equity implementations of any proposed beneficiary transfers by Suppliers should be carefully reviewed by DFID on a case by case basis.

12. Upside incentives should be included in future PBR programmes.

13. A substantial inception phase should be included in future PBR programmes, for putting monitoring and verification systems in place, where suppliers request this. This phase should be grant-funded, with no targets for output delivery.

14. The verification provider should be appointed before implementation begins, and programmes should be more prescriptive on monitoring, reporting and verification requirements.

15. Options for reducing the reporting and verification burden on Suppliers should be explored: for example by using mobile-based monitoring where suitable.

Table of contents

List of tables, figures and boxes

List of abbreviations

AiD	Associates in Development
AR	Annual Review
ASAL	Arid and semi-arid lands
ASWA	Accelerating Sanitation and Water for All
BCC	Behaviour change communication
CapManEx	Capital maintenance expenditure
CEO	Chief Executive Officer
CLTS	Community-led total sanitation
CV	Country Verifier
DEQs	Detailed evaluation questions
DFID	UK Department for International Development
DRC	Democratic Republic of Congo
FGD	Focus group discussion
HACT	Harmonised Approach to Cash Transfers
HEQs	High-level evaluation questions
HQ	Headquarters
INGO	International non-governmental organisation
ITA	Idara-e-Taleem-o-Aagahi
ITN	Invitation to negotiate
JMP	Joint Monitoring Programme
KII	Key informant interview
L&D	Learning and dissemination
LCBs	Local Capacity Builders
LOWASCO	Lodwar Water Supply Company
LPP	Lodhran Pilot Project
LQAS	Lot Quality Assurance Sampling
LV	Lead Verifier
M&E	Monitoring and evaluation
MDGs	Millennium Development Goals
MEAL	Monitoring, evaluation, accountability and learning

MER	Monitoring, evaluation and research
MV	Monitoring and verification
MVE	Monitoring, verification and evaluation
NCWSC	Nairobi City Water and Sewerage Company
NGO	Non-governmental organisation
ODF	Open defecation-free
ODI	Overseas Development Institute
OECD–DAC	Organisation for Economic Co-operation and Development Development Assistance Committee
OpEx	Operating expenditure
OPM	Oxford Policy Management
PACIDA	Pastoralist Community Initiative and Development Assistance
PATS	Pakistan Approaches to Total Sanitation
PBR	Payment by results
PEW	Programme Evaluation Workstream
PQQ	Pre-Qualification Questionnaire
QA	Quality assurance
RAG	Red, amber, green
RCT	Randomised control trial
RuFSAD	Rural Sanitation in Flood-affected Districts
SAFs	Sustainability Assessment Frameworks
SAWRP	South Asia WASH Results Programme
SNV	Netherlands Development Organisation
SSH4A	Sustainable Sanitation and Hygiene for All
SV	Shadow Verifier
SWIFT	Sustainable WASH in Fragile Contexts
UNICEF	UN Children's Fund
VEA	National 'Village et Ecole Assaini' programme in DRC
VFM	Value for money
WASH	Water, sanitation and hygiene
WEDC	Water Engineering and Development Centre
WSUP	Water and Sanitation for the Urban Poor

Key terms used in this document

The following table explains key terms used throughout this document.

Key term	Explanation
Consortium	There are three consortia, made up of groups of organisations, which have been contracted to implement the WASH Results Programme.
Country programme	Refers to activities at country level. There are 13 country programmes in the WASH Results Programme.
Country manager	The main contact person managing implementation in each country, within the three Supplier consortia.
Evaluation stakeholder	The organisations and people who are closely involved in the design of the evaluation and have a stake in its findings, namely DFID and the Suppliers.
Equitable	Refers to fairness. An equitable approach means taking account of the particular needs of different sub-groups within society. The outcome of an equitable approach is that all are included, regardless of gender, age (young and old), disability (including mental and physical disabilities), illness (including chronic illness such as HIV/AIDS), poverty, ethnic origin, caste, occupation, political affiliation, religion, sexual orientation, language or physical location. Those who tend to be neglected or marginalised, whether unconsciously or deliberately, are referred to in the following paragraphs as 'excluded' individuals or groups.
Impacts	Higher-level results, such as improved health, better educational results, increased income, and improved quality of life. Impacts describe the ultimate benefits enjoyed by WASH service users[1]. DFID's original theory of change only refers to health impacts—namely under-five mortality rates (per 1,000), and the proportion of the population affected by diarrhoeal diseases. In this document, we will clearly specify each time we refer to health or non-health impacts.
Lead Supplier	The organisation leading each of the three Supplier consortia.
LV	The three LVs lead the verification of each of the three Supplier contracts. These are employed by Itad within the e-Pact verification team.
MVE provider	e-Pact is providing the MVE under the WASH Results Programme. These activities are split between MV led by Itad, and evaluation services led by OPM.
Outcomes	The use of improved water supply and sanitation services; maintenance, care and financing of water supply and sanitation facilities; and the practice of good hygiene. In accordance with the WASH Results Programme logframe, 'outcomes' in this document refers to: the number of people using and continuing to use clean drinking water sources; the number of people using and continuing to use improved sanitation facilities; and the number of people hand-washing with soap and undertaking other hygienic practices at critical times, and continuing to do so after the output deadline in December 2015.
Outputs	Those physical and non-physical items which are delivered by WASH programmes, such as taps, toilets, newly established WASH committees, enhanced skills and knowledge. In accordance with the WASH Results Programme logframe, 'outputs' in this document refers to: the number of people gaining access to clean drinking water; the number of people gaining access to improved sanitation facilities; and the number of people reached through hygiene promotion activities through DFID support.

---

[1] This definition differs from the OECD's definition of 'impacts', which refer to any causal effects of the programme – at output-level, outcome-level, or 'impact'-level.

PBR	DFID uses PBR as an umbrella term for various performance-based contracts with different types of organisations. Payments to service providers such as the Suppliers implementing this programme are classed as 'results based financing' (DFID 2014a).
Payment milestones	Suppliers have committed to achieving specific deliverables at predefined dates, spread over the duration of their contracts. Suppliers are paid by DFID within a specified period, contingent on verification of the evidence submitted as part of the payment milestone.
Programme level	Refers to evaluation activities or questions which relate to the whole WASH Results Programme.
Service users	The target population that is intended to benefit from the WASH Results Programme's activities.
Sub-country programme	Refers to implementation activities carried out by a sub-set of local partners, within a country programme. There are two to five local partners within each country programme.
Supplier	The organisations contracted by DFID to implement the WASH Result Programme.
Sustainable	Sustainability relates to whether WASH services and good hygiene practices continue to work and deliver benefits over time, thus delivering permanent beneficial change in WASH services and hygiene practices (Abrams, no date; WaterAid, 2011). This is determined by a variety of factors. In this evaluation, we will therefore approach sustainability in two steps: first, as part of the mid-term evaluation in 2016, we adopt a forward-looking (risk-based) approach, examining the factors which determine the likelihood of outputs and outcomes being sustained until 2018 and beyond. Secondly, as part of the endline evaluation in 2018, we will examine if and why outputs and outcomes were sustained until 2018, assessing as well their likelihood of being sustained beyond the lifetime of the programme.
Target audience of the evaluation	The organisations and people to whom the findings of the evaluation will be disseminated. This includes the evaluation stakeholders, other donors with an interest in PBR, WASH practitioners and the evaluation community.
Unsuccessful bidder	The organisations which expressed an interest in the WASH Results Programme, but were not selected to implement it.

1   Introduction

1.1   Evaluation objectives

This evaluation falls under the MVE component of the WASH Results Programme, implemented by the e-Pact consortium under DFID's Global Evaluations Framework Agreement. The evaluation component is led by OPM, and the MV component is led by Itad.

The evaluation is primarily being undertaken **for learning purposes,** not as an accountability exercise. Its objectives are to assess:

(i)   whether the programme successfully achieved its stated objectives;
(ii)   the influence of programme design, including the PBR modality, on this achievement; and
(iii)   lessons for applying PBR in WASH programming going forward.

These objectives are being be met by answering a sub-set of the six high-level evaluation questions (HEQs) in this mid-term evaluation report, and the remainder in the Endline Evaluation Report in 2018. Each HEQ is broken down into a number of Detailed Evaluation Questions DEQs; see Section 3.4.

The Evaluation Design Document, submitted in May 2015, set out the comprehensive design of the evaluation: our overarching evaluation approach, and our specific methodology for the various components of the evaluation. The design was informed by DFID's Evaluation Framework for Payment by Results (2015), which lays particular emphasis on the effect of PBR on supplier behaviour.

In pursuing these objectives, the evaluation has also identified lessons and recommendations that should be of use to the Suppliers and to DFID for the remaining implementation period. For this reason, this evaluation also has a substantial **formative evaluation component**. The report should also be of interest, however, to other donors and organisations working on the delivery of WASH results at scale, and to the WASH sector in general.

1.2   Scope of this report

This mid-term evaluation report is the first evaluation output under the MVE assignment. It examines progress half way through the four-year WASH Results Programme. It answers the sub-set of evaluation questions which relate to the original design and tendering process of the DFID-funded WASH Results Programme, and to its Output Phase (March 2014–December 2015, later extended to March 2016) (see Section 3.4 for detail on which evaluation aspects are covered). The remainder of the evaluation aspects will be covered by the endline evaluation report in 2018, at the end of the programme.

This report comes under the PEW, which provides the assessment of overall programme performance. The findings of the other workstream, the RCT Research Study, which examines factors affecting the sustainability of sanitation behaviour change in Pakistan irrespective of PBR, are described in a separate RCT Baseline Report (May 2016).

1.3   Structure of this document

The remainder of this report is structured as follows:

Chapter 2 presents the background to the WASH Results Programme.

Chapter 3 outlines the evaluation approach.

Chapter 4 reviews the overall design of the programme by DFID, and compares this to the design of the three Supplier programmes, each implemented by a different consortium or organisation. Together, these make up the overall programme.

Chapter 5 describes the design of the verification component.

Chapter 6 reviews Output Phase results across the three Suppliers, drawing on three country case studies (one for each Supplier) as well as documentation reviews and wider discussions held with DFID, the MV team and the Supplier staff at headquarters (HQ).

Chapter 7 provides a forward view in terms of Outcome Phase plans, and the prospects for equity and sustainability.

Chapters 8 and 9 set out the evaluation conclusions and recommendations.

Additional detail is provided in the annexes:

- **Annex A** contains terms of reference for the MVE component of the programme (including this evaluation).
- **Annex B** critically reviews the programme theory of change, as defined in the Business Case.
- **Annex C** presents the full evaluation matrix and the headline evaluation findings alongside each evaluation question, as well as a list of respondents interviewed.
- **Annex D** contains the three country case studies and a short evaluation report on SSH4A in countries beyond Uganda.

2   Background to the WASH Results Programme results

This chapter provides background information on the emergence of the WASH Results Programme. The detailed design of the programme is examined in Chapter 4.

Key findings of this chapter
• The concept of the WASH Results Programme was born out of the UK's pledge to contribute to global MDG commitments by ensuring that 60 million additional people gained access to WASH services by December 2015.
• The programme underwent a long design phase during 2012 and 2013, during which extensive consultations were held on the scale, focus and financing modality of the programme. Up to five contracts, with a total value of £109 million, were envisaged in the Business Case, benefitting 7.1 million people by December 2015 and up to 9.8 million people by 2018. Following a competitive tender process, three contracts were awarded, with a total value of £75 million, targeting 4.5 million beneficiaries by December 2015 and more by programme end. Suppliers recall that the PBR modality was confirmed relatively late in the design phase (July 2013).
• The competitive tender process lasted from September 2013 to January 2014, with contract negotiations lasting until March 2014. Bidders found the tendering process challenging, not only because PBR was new and unfamiliar, but also because the invitation to negotiate (ITN) lacked clarity on several key points, which were only resolved after contracting (a consequence of using a negotiated approach).
• The period for the delivery of outputs was shortened from approximately three years to around 18 months (June 2014 to December 2015).

2.1   The use of RBF in international development

In June 2014, DFID published 'Sharpening incentives to perform: DFID's strategy for payment by results'. This made a commitment to greater use of PBR for the funding of international development programmes. More recently, the 2015 Conservative Party Manifesto committed the new UK government to increasing the use of PBR across all departments. Box 1 below gives an overview of DFID's definitions of PBR.

Evidence on the effectiveness of PBR in international development programmes is currently limited. DFID has acknowledged this and signalled the need for more learning on what works in PBR, and under what circumstances. The use of PBR in the WASH sector specifically is relatively new and the available literature is limited, though growing.[2] Where PBR has been applied, this has more commonly been in the form of results-based aid to a government agency; its use for the contracting of INGOs to deliver community-based WASH projects at scale – as is the case under this programme – is rare.

---

[2] See, for example, www.gpoba.org/node/787.

Box 1: DFID's definitions of PBR

DFID uses PBR as an umbrella term for various performance-based contracts entered into with different types of organisations and has identified three broad typologies of PBR based on the type of organisation that (some of) the risk for the delivery of results is transferred to:

- results-based aid, in which results payments are made to partner governments;

- results-based financing, in which results payments are made to service providers/suppliers (contracted directly or indirectly via an intermediary); and

- development impact bonds, whereby investors bear some (or all) of the risk of the implementation and delivery of results by putting money upfront into a development programme and are then paid upon the delivery improved results.

Within these three broad categories, the use of PBR varies in two areas:

- the level of payments on delivery (upfront/ex-post): the WASH Results Programme entails 100% of payments *ex-post*;

- the level of result the payment is linked to (outcomes, outputs or processes): under the WASH Results Programme, payments are made for processes and outputs and outcomes.

2.2   Origins of the WASH Results Programme

2.2.1   DFID's WASH commitments

The concept of the WASH Results Programme was born out of the UK's pledge, made by the Secretary of State for International Development, Andrew Mitchell, in April 2012, to contribute to global MDG commitments by ensuring that 60 million additional people gained access to WASH services by December 2015 (DFID 2014c). At the time the commitment was made, decisions had not yet been made on how they would be delivered.

> *Several options were considered for reaching the UK's MDG commitment. There was a clear preference for national programmes tailored to national contexts, designed together with country WASH advisers. A large centrally managed programme was a 'last resort'.* – DFID WASH adviser.

2.2.2   Finding large-scale partners

A secondary, but important, driver of the programme was DFID's desire to expand its base of potential large-scale WASH providers. Prior to the WASH Results Programme, DFID had one major non-government implementing partner (the UN Children's Fund (UNICEF)) with the capacity to absorb and utilise large investments.

> *An explicit (but unwritten) goal of the programme was to find other large-scale suppliers for future programming.* – DFID WASH Team Leader.

2.2.3   Development of the programme, 2012 to 2014

DFID initially approached several of their country offices in 2012 to establish whether they could scale up their WASH programmes to meet the Secretary of State's commitment. At the same time a new grant was negotiated, providing support to UNICEF through the nine-country 'Accelerating Sanitation and Water for All' (ASWA) programme. ASWA began in July 2013 and was due to finish

---

in September 2016; however, an extension to deliver additional results linked to a new Government target, up to 2018, has recently been approved.

A third action by DFID involved an exploration of the appetite and capability of INGOs and private firms to implement large-scale WASH work through a results-based programme. To this end an Early Market Engagement event took place in September 2012, attended by more than 70 organisations. At this early stage, some WASH professionals expressed serious concerns about an over-emphasis on beneficiary numbers and target dates, potentially at the expense of quality and sustainability of services and behaviour changes. DFID concluded that there was sufficient interest and enough competition to warrant a competitive tender process and it also acknowledged the concerns voiced about quality and sustainability.

Confirming projections for the DFID country programmes took longer than expected but once commitments were secured, a decision was made to go ahead with the WASH Results Programme to cover the estimated shortfall in beneficiary numbers needed to meet the UK's MDG commitment. In June 2013, DFID was still describing the programme as a five-year 'Results Challenge Fund' targeting 9.8 million people by programme end, with several different funding modalities considered in the Business Case. A month later, in July 2013, a second Early Market Engagement meeting was held with potential bidders, announcing a budget of £109 million for up to five 'large results contracts'. By this stage in the process DFID was talking of 'significant levels of PBR'. The Suppliers we interviewed began choosing implementation partners before the PBR element was announced.

By August 2013 the Business Case had been approved by the Secretary of State, and the tendering process was agreed. The Pre-Qualification Questionnaire (PQQ) was issued on 28 August, with the response deadline set as 26 September 2013. The final terms of reference were for three contracts, to serve 4.5 million beneficiaries by December 2015, with a total budgetary envelope of £75 million. The reason for the reduced budget envelope was that it had become clear that the UK Government target would almost be met by a combination of DFID country programmes, existing centrally managed programmes and core contributions to multilateral agencies, and thus there was only a relatively small shortfall to make up.

The final stage for pre-qualified bidders, the ITN (full proposal stage), commenced in October 2013, although forms and guidance were not issued until mid-November. Submission of full proposals was required by 05 December 2013. Out of 21 expressions of interest, 10 suppliers were invited to submit proposals. Three of these subsequently withdrew, leaving a total of seven proposals.

Bidders have commented that the tendering process was challenging, not only because the programme design was new and unfamiliar, but also because the ITN lacked clarity on some key points. As a result, a large number of questions were submitted to DFID, who had to issue numerous rounds of clarifications. The programme design was in fact new for both DFID and potential suppliers, and some of the detailed content – especially how some aspects of the PBR mechanism would work in practice – was yet to be worked out. The level of ambition DFID was looking for at outcome level was also not clear as the logframe would only be finalised after suppliers had been appointed. There was, therefore, considerable uncertainty on both sides at this stage of the process. The procurement process took several months; though it should be noted that longer procurement times are common under an OJEU-negotiated procurement process.

The final selection of preferred bidders (the 'Suppliers') was announced 23 January 2014, contract negotiations took place (between January and April), and contracts were issued. Three Supplier contracts, totalling £75.5 million, were issued for a 48-month programme (1 April 2014 to 31 March 2018). A separate MVE contract, worth £3.46 million and spanning 54-months (May 2014 to

November 2018). The MVE contract was agreed a few weeks later than the Supplier contracts, a point which had subsequent implications for verification, in particular.

The chronology of the events described above is presented in Figure 1. By the time the programme commenced in April / May 2014, the contracts awarded differed from the ceiling set out in the Business Case:

- the overall programme duration was four years (cf. the five-year ceiling in the Business Case);
- three contracts were awarded, targeting a total of 7.5 million beneficiaries, with a total value of £75 million. The approved Business Case made provision for up to five contracts, targeting 9.8 million people, with a total budget of £109 million; and
- the time-scale for delivery of outputs to beneficiaries had been reduced from approximately three years (the initial expectation) to around 18 months (from commencement of contracts to the MDGs deadline of December 2015). This period was, however, followed by a further two-year phase focusing on outcomes and sustainability.

The overarching programme features designed by DFID, and translated by Suppliers into various country programmes, are described and analysed in Chapter 4.

Figure 1:   Timeline of key design features of the WASH Results Programme
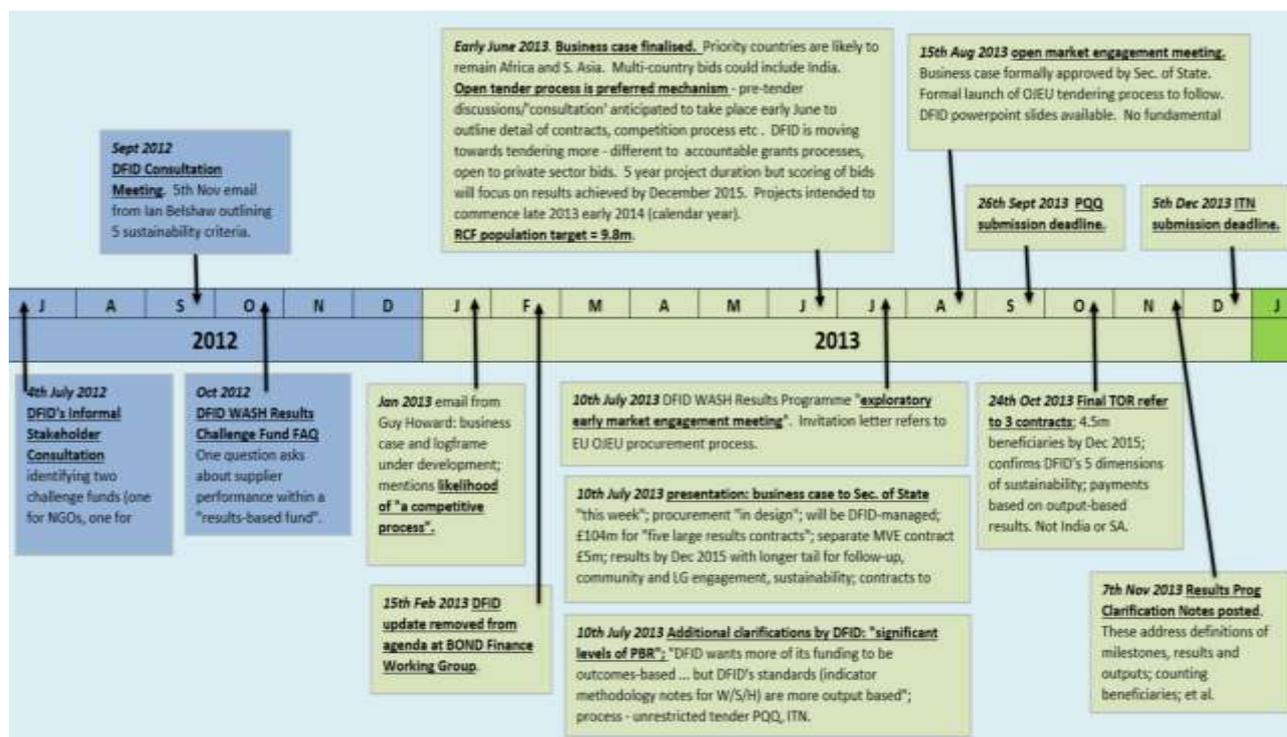
3   Evaluation approach

Key findings of this chapter

- Evaluation instruments are applied at programme level and at country level, in order to capture lessons that are relevant both for the whole programme and for specific case study countries.

- Programme-level evaluation activities included reviewing programme documentation and interviewing DFID, the MV team and Lead Supplier staff at HQ level.

- Evaluation activities at country level include three case studies and an RCT in Pakistan. This report does not cover the RCT, which is being documented in a separate baseline report (May 2015).

- Country-level evaluation activities included reviewing programme documentation and interviewing Supplier staff, communities and national WASH stakeholders during visits to Kenya, Pakistan and Uganda. Each case study country was visited in early 2016 and will be visited again in early 2018.

- Evidence synthesis is being undertaken in mid-2016 (this report) and in mid-2018, to draw together findings across the programme at the end of the Output and Outcome Phases, respectively.

3.1   Current programme status

At the time of writing, the three Suppliers have completed delivery of their contracted outputs and DFID's target result of 4.5 million unique beneficiaries of water supply, sanitation and hygiene improvements has been delivered and independently verified[3]. Brief details of the three programmes are provided in Table 1 below, to provide some context for the evaluation approach. The detailed content of each programme is examined further in Section 4.5.

Table 1:   The Supplier programmes

Programme	Lead Supplier and consortium partners	Countries of operation	Programme components
SAWRP	Plan International, with WaterAid, Water Sanitation for the Urban Poor (WSUP) and Unilever. Water Engineering and Development Centre (WEDC) and Ipsos provide support on monitoring and knowledge management	Pakistan, Bangladesh	Sanitation; hygiene, including a substantial school hygiene promotion component; and a modest water supply component. (All rural)
SSH4A	Netherlands Development Organisation (SNV) only (not a consortium)	Ghana, Ethiopia, Tanzania, Uganda, Kenya, South Sudan, Mozambique, Zambia, Nepal	Rural sanitation and hygiene only
SWIFT	Oxfam, with Tear Fund, WSUP, Practical Action, Sanergy, Concern Worldwide. The Overseas Development Institute (ODI) supports monitoring, evaluation, accountability and learning (MEAL)	Democratic Republic of Congo (DRC), Kenya. (Also Liberia originally, but cancelled due to the Ebola crisis)	Water supply, sanitation and hygiene. (Urban and rural)

---

[3] This review does not question the accuracy of the reported numbers, though we do have comments on the basis for counting results.

3.2   Overview

The evaluation of the WASH Results Programme involves both an assessment of overall programme performance (PEW), and an RCT Research Study in Pakistan[4]. The evaluation methodology was designed and refined during the Inception Phase and the Exploratory Phase, while the bulk of data collection is taking place in the ongoing Explanatory Phase (Q3 2015 – Q2 2018).

This Mid-Line Evaluation Report falls under the PEW, which provides the assessment of overall programme performance against a sub-set of the evaluation questions, with the remainder being addressed in the endline evaluation report in 2018. Detail on which sub-set of evaluation questions will be addressed in this report is outlined in Section 3.4.

Section 3.3 below explains in detail the sources of evidence which were used for this evaluation report. These included information capture at **programme level and at country level,** in order to capture lessons relevant both to the programme overall and to specific case study countries (Kenya, Pakistan and Uganda).

Evidence synthesis is being undertaken in mid-2016 (this report) and in mid-2018, to draw together findings across the programme at the end of the Output and Outcome Phases, respectively.

Evaluation findings are being disseminated to DFID and implementing partners through annual Supplier learning workshops, and to the wider sector via other workshops and reports. Dissemination activities are managed by the Learning and Dissemination (L&D) team within Itad.

3.3   Sources of evidence used

This report draws on evidence from a number of evaluation activities, which are set out in the sub-sections below. These match the activities set out in the Evaluation Design Document (May 2015), with the exception of the following:

- **the in-depth value for money (VFM) analysis** has been postponed until 2017 because more detailed financial data are required from Suppliers; and
- **the community-level key informant interviews (KIIs) and focus group discussions (FGDs)** have been postponed to 2018, and will be implemented as a more in-depth 'beneficiary impact assessment', to be carried out at endline. In 2016, only a limited number of days were spent at the community level during the country visits, primarily to provide some context to the evaluation activities. For budgetary reasons, it was decided to allocate the bulk of community-level resources to the endline, to allow a detailed assessment to take place then, rather than spreading resources across two light-touch assessments.

The summary table in Section 3.7 outlines exactly which data sources were used to answer which evaluation question. The full list of people interviewed as part of this report is set out in Annex C.2.

---

[4] The findings of the RCT Research Study, which examines factors affecting the sustainability of sanitation behaviour change in Pakistan irrespective of the PBR modality, are described in a separate RCT Baseline Report (May 2016).

3.3.1 Programme-level evaluation activities

A combination of documentary reviews and interviews was carried out between September 2014 and December 2015, including the following:

- **Literature review of best practice in the application of PBR,** not only in the WASH sector but within the broader context of donor-funded programmes implemented by INGOs. This confirmed that 100% PBR, as applied in the WASH Results Programme, is rare; its use in the WASH sector generally is uncommon; and its application to INGO Suppliers (as opposed to government recipients) is also unusual.

- **Review of DFID's programme documentation** (the Business Case, tender documentation and background reports) to understand how the programme was designed. Interviews were also held with the key DFID staff involved in the design of the WASH Results Programme.

- **Review of Supplier documentation and discussions with each Supplier**, to understand the organisational structure of each consortium; their operational approach, key activities and targets; and the timeline for implementation. The review included reviewing the contracts of each consortium, to clarify how the PBR model was designed for their programme. We identified features which were common across all PBR contracts, as well as any significant differences between contracts, and considered their implications for meeting output targets. Quarterly remote catch-ups were held with each Lead Supplier from September 2014 to December 2015, to understand how activities were being implemented in each consortium and to understand any issues that arose.

- **Interviews with bidders who withdrew or were unsuccessful during the tendering process,** exploring what initially attracted them to the WASH Results Programme, any doubts or concerns they had, and what reasons (where applicable) led them to withdraw.

- **Interviews with DFID** explored the programme design process and factors which influenced it (March 2015), as well as exploring the factors that affected programme performance against output targets (February 2016).

- **Quarterly catch-ups were held with each Lead Supplier** throughout September 2014 and December 2015, to understand how the implementation activities were progressing. Detailed discussions were also held on this at the **annual learning events** with Supplier headquarters staff, MV team members and (on occasion) DFID UK staff, which took place in March 2015 and March 2016 in London. These provided valuable insights into how PBR was working out in practice – particularly the verification process – and an opportunity to consider whether and how programme design could be improved for future interventions.

- After completion of each case study visit (see Section 3.3.3), telephone discussions took place in June/July 2016 with some of the Supplier country managers of the countries not visited, to understand how experiences differed in non-case study countries. Phone interviews were carried out with the SWIFT DRC country manager, and separately with the Lead of the SAWRP consortium (as the Bangladesh Country Manager was not available). A two-hour telephone conference was held with SNV country managers from Ethiopia, Ghana, Kenya, Mozambique, South Sudan, Tanzania, Zambia and Uganda (the Nepal country manager commented in writing due to time zone issues). These interviews explored if and why those factors which were identified as affecting implementation in the three case study countries visited also affected implementation in the 10 remaining countries.

- **Analysis of Supplier monitoring data** was undertaken to assess implementation progress, the extent of equitable results (where feasible) and any reductions in payments. As all Supplier monitoring data were subject to very rigorous third-party verification by the MV team, the evaluation team did not undertake any additional quality assurance of the secondary monitoring data and considered them to be accurate.

- **Review of verification reports and discussions with the MV team**, to understand the design and progress of the verification activities. **Quarterly catch-ups** were held with each LV throughout September 2014 and September 2016, to understand how the verification process was being implemented and to understand any issues that arose. This catch-up included a review of every quarterly verification report.
- **Interviews with the Itad L&D team,** to understand what learning has taken place during the programme and how this learning is feeding into better evidence-based WASH policy and programming.

Table 2: Timeline of programme-level activities

The green dots indicate the completed activities that provide the evidence feeding into this Mid-Line Evaluation Report. The remainder (black dots) will inform the 2018 Endline Evaluation Report.

	2014	2015				2016				2017				2018			
	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Documentation review	🟢	🟢															
Interviews with unsuccessful bidders	🟢																
Interviews with DFID staff		🟢					🟢								●		●
Interviews with Lead Suppliers	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢		●		●		●		●
Remote interviews with country managers							🟢										●
Remote interviews with DFID in-country staff																	●
Analysis of Supplier monitoring data		🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢		●		●		●		●
Interviews with the MV team		🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢		●		●		●		●
Interviews with L&D team							🟢								●		●

Note: The blue shading indicates the timing of country-level data collection (see Section 3.3.3).

3.3.2 Selection of one country case study for each Supplier

Supplementing the programme-level activities set out above, in-depth country case studies are being carried out: case studies took place at the end of the Output Phase in early 2016 and will take place again at the end of the Outcome Phase in 2018. In 2016, questions focused on the experience of the Suppliers and service users during the Output Phase, focusing in particular on the factors driving efficiency and effectiveness. In 2018, questions will focus more on potential impact and the likelihood of long-term sustainability.

The country case studies provide an in-depth understanding of Supplier behaviour at country level, in terms of how the programme is implemented and delivers its outputs. This includes exploring the design choices made to meet WASH targets in country, how country-specific factors influenced such choices, and under what circumstances key programme features, particularly the PBR modality, influenced country-level Supplier behaviour and intervention implementation. Such analysis enables investigation of the causal mechanisms that explain the translation of activities into outputs, as postulated in the theory of the change of the programme.

Case study countries were selected purposively and as such are not statistically representative of the programme as a whole. The primary basis for selection was to cover a range of scenarios which relate back to the aspects explored in the evaluation questions: selecting cases with variations in terms of organisational complexity of the Supplier and the range of programme activities allowed evaluation questions on the effectiveness of the PBR modality under different circumstanced to be explored in detail. Due to budgetary constraints, only three case studies were feasible (see Section 3.6). The following case study countries were selected and agreed with DFID following sign-off of the Evaluation Design Document:

- **Kenya (SWIFT consortium)** is an interesting case for understanding how WASH activities and the PBR modality were implemented within a consortium that includes very diverse partners and implementation activities;

- **Uganda (SSH4A programme)** highlights the opposite case, allowing the evaluation team to explore how WASH activities and the PBR modality were implemented by a single INGO applying a more or less standard implementation model across multiple countries; and

- **Pakistan (SAWRP consortium)** is implemented by three principal partners. Two of these apply a broadly similar approach to sanitation and hygiene promotion, while the third delivers a dedicated school hygiene promotion intervention with a unique methodology.

3.3.3 Evaluation activities in country

In order to understand the circumstances which have shaped the achievements of each case study country programme (and the influence of PBR on this achievement) in-depth investigations were undertaken to provide a clearer understanding of the sequence of events and the motivations of key actors. By the end of the evaluation each case study country will have been visited twice, first in early 2016 (for this report) and again in early 2018 (for the endline report).

Country visits were scheduled over two weeks, entailing the following activities (each evaluation team included a consultant from the country involved):

- The initial three to five days in the capital city were used to hold group discussions and KIIs with **Supplier staff and staff of the local partners**, and to review relevant documentation and data. This included discussions on how and why activities were designed in the chosen manner, and how key programme features, including the PBR modality, had affected implementation to date.

- This time was also used, where possible, to hold **discussions with sector experts** (including DFID staff) not directly involved in the programme, to better understand the wider WASH context within which programme activities were being implemented. Interviews were also held with **relevant government staff**.

- Light-touch field visits were made to a handful of communities where a sub-set of local partners had intervened. Typically, these visits covered two to three local partners in different geographical regions, and included discussions with beneficiary communities wherever possible. **Group discussions and interviews were held with community leaders and community members**.

- Following the field visits, supplementary interviews and discussions were held with government staff and Supplier staff in the capital city, ending with a de-briefing session with Suppliers and their partners.

The findings from the case studies were compared to experiences in the remaining countries via **telephone discussions held with Supplier country managers of the countries not visited** (see Section 3.3.1). In reviewing findings from the country case studies, we have been careful to

distinguish between findings which are generalisable across the whole programme and those which are interesting but specific to particular projects, implementing partners or locations.

Table 3:   Timeline of different evaluation activities in case study countries

The green dots indicate the completed activities that provide the evidence feeding into this Mid-Line Evaluation Report. The remainder (black dots) will inform the 2018 Endline Evaluation Report.

	2014	2015				2016				2017				2018			
	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Supplier country staff						●								●			
KII with national sector stakeholders						●								●			
KII with local implementing partners						●								●			
KII with community leaders														●			
FGD with community members														●			

3.3.4   VFM analysis

As part of the overall evaluation of the WASH Results Programme, the evaluation is exploring the VFM of programme activities. Rather than simply assessing *whether* funds were spent in a cost-efficient and cost-effective manner, the VFM analysis explores *why* VFM varied between different countries, and explores drivers of VFM under a PBR modality.

A light-touch analysis of price-per-beneficiary was undertaken in each case study country at the end of the Output Phase (March 2016). This analysis will be repeated in 2018. Analysis was based on the limited financial and monitoring data shared by Suppliers, as well as qualitative interviews with country teams during the case study visit.

Only light-touch analysis of price-per-beneficiary was feasible at this stage due to the sensitivities around sharing financial data under a PBR programme. A more in-depth analysis of cost-per-beneficiary is planned for 2017, pending confirmation from DFID.

3.4   Evaluation questions

3.4.1   Evaluation aspects covered by the whole evaluation

While the evaluation examines **what** was achieved, the main focus is on **how** and **why** and **under what circumstances** results were achieved. In short, it sets out to answer the following questions:

- What were the **expectations** of the programme, and how did these change as the programme went along?
- To what extent were the **theories of change** and the **designs** set out by DFID and the Suppliers, and the **implementation approaches** taken, likely to lead to high quality, sustainable and equitable results?
- Given the PBR modality, to what extent were the **MV** systems and processes fit-for-purpose?

- What were the **achievements** of the programme (both those which were verified and paid for, and those outside the scope of the contract results packages)?
- What are the **prospects for the sustainability and equity** of outcomes beyond the end of the programme?
- What are the **prospects for health and non-health impacts** beyond the end of the programme?
- What **recommendations** can be made, based on the evidence generated, for those designing large-scale WASH interventions in future?

The full set of evaluation questions is set out in Annex C1. It includes the data sources and methodology to be used to answer each question and summarises the evaluability assessment.

3.4.2   Evaluation aspects covered by this report

This report focuses on a sub-set of evaluation questions (see Table 4): namely, **relevance**, **efficiency** and those aspects of **effectiveness** which relate to outputs. The likelihood of the **impacts** of programme results being achieved and **sustained** are only briefly touched upon in this report and will be covered in detail in the endline evaluation report in 2018. The **effectiveness** of the programme with regard to outcomes will also only be covered in 2018. In addition, a separate in-depth VFM analysis planned for 2017 will explore aspects of programme efficiency in more depth.

Table 4:   DEQs explored in this report

HEQs	DEQs
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?
	DEQ 1.2: To what extent did the **programme's design** (i.e. the theory of change) 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 in regard to achieving the 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's **'learning objectives'**?
	DEQ 1.9. To what extent was the **design of each consortium sub-programme** appropriate for achieving DFID's key objectives?
HEQ2 Effectiveness: To	DEQ 2.1: To what extent did the programme achieve the intended outputs at scale?

what extent did the programme operate as intended, and which factors helped/hindered the achievement of output and outcome objectives (under what circumstances)?	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 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within **consortia sub-programmes**?
	DEQ 2.5: Under what circumstances did the **PBR framework** help/hinder the achievement of intended outputs and outcomes?
	DEQ 2.6: Under what circumstances did the **PBR framework** affect the quality of programme implementation (positive or negative)?
	DEQ 2.7: Under what circumstances did Suppliers implement innovative approaches and focus on learning?
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?
	DEQ 3.2: Under what circumstances did the **PBR modality** affect the cost-effectiveness of individual sub-programmes?
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**?
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?

3.5   Evaluation method used

The analysis underlying this evaluation report is based on the methods outlined below. The likelihood of the **impacts** of programme results being achieved and **sustained** are only briefly touched upon in this report – based on likely determining factors set out in the Evaluation Design Document (May 2015) – and will be covered in detail in the endline evaluation report in 2018.

Detail on the basis for the evaluative judgement made for each evaluation question is given in a column in Table 15 in Section 8.1, alongside a statement of the confidence with which the evaluation assessment is made.

3.5.1   Assessing the relevance of the WASH Results Programme

Making an evaluative judgement on the relevance of the theory of change required a critical analysis of the assumptions underlying it, in order to determine the strengths and weaknesses of the programme design. Analysis first involved assessing whether the programme's objectives were valid and were appropriate in regard to meeting sector needs, then reviewing whether the elements of the theory of change were appropriate in regard to achieving programme objectives.

We based our assessment both on the examination of secondary sources as well as primary data, in the form of interviews with the DFID and key programme stakeholders. To obtain a balanced picture of how the WASH Results Programme was designed, we not only examined the theory of change itself but also the Programme Business Case. A balanced perspective was obtained by not only interviewing suppliers that were successful but also those bidders that were unsuccessful during the tender process (see Section 3.3.1 for detail).

3.5.2   Preliminary assessment of efficiency

Making an evaluative judgement on the efficiency of the WASH Results Programme was not based directly on the theory of change, but relied on the conceptual framework, which outlines the elements which drive efficiency (see Design Document). This framework allowed us to make an evaluative judgement about the efficiency of the WASH Results Programme based on the presence or absence of these elements.

Assessing *if* efficiency has been achieved can be directly measured through VFM analysis, based on an analysis that compares the outputs achieved with the programme spending on each output (secondary data). However, as has been stated, due to the sensitivities around sharing financial data under a PBR programme, only an analysis of *price*-per-beneficiary could be undertaken at this stage. A more in-depth analysis of *cost*-per-beneficiary is planned for 2017, pending confirmation from DFID.

A preliminary analysis of *which factors* caused any observed changes in efficiency was carried out by comparing the perspectives of Supplier staff at country level and at programme level (Lead Suppliers). A more in-depth analysis of these factors will be undertaken as part of the cost-per-beneficiary analysis planned for 2017.

3.5.3   A theory-based assessment of effectiveness (outputs only)

The evaluation uses a **theory-based approach** to explore the effectiveness of the programme to date. This approach entails using the theory of change of the programme as a starting point to test the assumptions implicit in DFID's original programme design. In other words, we explore (a) whether the programme could reasonably be expected to deliver its intended results given the way it was designed and structured, and (b) the extent to which it did, or did not, deliver the hoped-for outputs, and the reasons why. This report only focuses on the factors affecting outputs, whereas the endline evaluation in 2018 will also include factors affecting outcomes. The evaluation team based its evaluation approach on two theories of change: first, DFID's programme-wide theory of change (see Annex B1), and, secondly, a second theory of change (see Annex B2) constructed by the evaluation team to provide further detail on the underlying PBR modality.

The evaluation activities included gathering detailed information on the causal factors and contextual factors affecting output results. Further detail on these factors will be gathered for the endline evaluation report, where all factors will be written up under the contribution analysis

framework. Contribution analysis draws causal inferences by explaining **how change is generated**, based on a plausible and verified theory of change, and recognising that the programme is unlikely to be the sole factor producing change – rather, it produces change in interaction with contextual factors. A detailed explanation of contribution analysis was provided in the Evaluation Design Document.

3.6   Methodological limitations

The evaluability assessment carried out as part of the Evaluation Design (May 2015) set the boundaries of evaluation to only include those evaluation aspects which can feasibly be addressed under the evaluation timeline and with the available resources.

As it was not possible to construct a counterfactual for the PBR modality, one of the biggest limitations of the evaluation will be isolating how the PBR modality has affected programme performance from the effect of other programme and non-programme factors. While we have designed an approach that aims to minimise any bias in opinions regarding PBR effectiveness by Supplier staff, by comparing their views to non-programme staff in the country case studies, it is more challenging to compare views for countries that did not form part of the case studies, as there may still be a limited range of people to draw on to obtain an external view of the programme. To highlight potential limitations, we clearly present who was interviewed.

The Lead Suppliers, country managers and the LVs were all given the opportunity to comment on the draft version of this report two weeks before it was submitted to DFID. This allowed findings to be further cross-checked.

All stakeholders have been given the opportunity to comment on the draft findings.

In addition, certain limitations remain, relating to the scope of the evaluation. The evaluation aims to pull together lessons on WASH implementation and PBR design which are relevant to the wide variety of contexts present across the 13 country programmes of the WASH Results Programme. While we have designed a rigorous approach for pulling together findings and checking their validity across the WASH Results Programme, we are aware that not all findings from the three case studies will be generalisable to the overall 13 country programmes. Therefore, when summarising findings from the case studies, we will clearly indicate which of these are generalisable, and which are interesting but not widely applicable across the programme. Budget constraints do not permit us to undertake more than three country case studies.

Budgetary constraints also led to the decision to allocate the bulk of community-level resources to the endline, to allow a detailed assessment to take place then, rather than spreading resources across two light-touch assessments at mid-line and endline. As a result, this report only provides very limited insights into community-level aspects, which will be picked up in more detail in 2018.

3.7   Summary table of data sources for the evaluation questions addressed in this report

The table below gives the data sources and methodology used to answer those evaluation questions addressed in this report. The full table covering also those evaluation questions to be addressed in 2018 is available in Annex C1.

The 'relevance' aspects of the evaluation are covered by Chapters 4 and 5, efficiency and effectiveness are covered by Chapter 6, and forward-looking reflections on likely sustainability and impact are predominantly covered by Chapter 7.

Table 5:   Data sources for the evaluation questions addressed in this report

DEQs	Data sources	Methodology / basis for evaluation judgement	Report section
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?	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation • WASH best practice review	Evidence of clear articulation of all of DFID's underlying programme objectives in the Business Case, the ITT and other DFID documentation. This was triangulated with the views of DFID, Suppliers, and unsuccessful bidders. Consistency of stated objectives between documents.	4.1, 4.2, 4.3
DEQ 1.2: To what extent does the programme's design (i.e. the theory of change) set out a clear and realistic process for how programme activities will achieve intended outputs, outcomes and impacts?	**Primary** • DFID interviews • Lead Supplier interviews **Secondary** • DFID documentation	Evidence of consistency between DFID interviews and DFID documentation, triangulated with the views of DFID and Suppliers. Evidence of completeness of theory of change assumptions in DFID documentation.	4.4
DEQ 1.3: To what extent were the scale and pace of the programme (including the Dec. 2015 deadline) realistic in regard to achieving intended outputs and outcomes, given the capacity of Suppliers and their local partners?	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation	Evidence in the ITT of tailoring of WASH Results Programme design to capacity and context, triangulated with the views of DFID, Suppliers, and unsuccessful bidders.	4.6
DEQ 1.4: To what extent was the PBR modality appropriate in regard to achieving sustainable and inclusive WASH outcomes, given the capacity of Suppliers and the timeline of the programme?	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews • LV interviews **Secondary** • DFID documentation • Verification Design Document	Evidence of the expected causal link between PBR and achieving the objectives in programme documentation and verification methodology, triangulated with the views of DFID, Suppliers, Verifiers and unsuccessful bidders. Perceptions of whether there is sufficient capacity in DFID and expected bidders to implement a PBR programme (triangulated across different interviewees).	4.7

DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation • WASH best practice review	Evidence of clear intent of the programme to encourage 'innovative' private sector partnerships, based on DFID documentation. Assessment of likelihood of private sector partnerships based on triangulation of perceptions of DFID, Suppliers, and unsuccessful bidders.	4.8.1
DEQ 1.6: How likely was it that the programme design would encourage Suppliers to propose 'innovative WASH interventions'?	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation	Evidence of clear intent of the programme to encourage 'innovative' WASH interventions, inclusive results and learning, based on DFID documentation. Assessment of the extent to which each was achieved by the end of the Output Phase, based on triangulation of perceptions of DFID, Suppliers, and unsuccessful bidders.	4.8.1
DEQ 1.7: How likely was it that the programme design would encourage inclusive outputs and outcomes?			4.8.3
DEQ 1.8: How appropriate was the WASH Results Programme's design in regard to achieving the programme 'learning objectives'?			6.5
DEQ 1.9. To what extent was the design of each consortium sub-programme appropriate in regard to achieving DFID's key objectives?	**Primary** • Lead Supplier interviews • Supplier country case studies • Country manager survey **Secondary** • Quarterly verification reports	Evidence of consistency between the objectives of DFID's design and the design of the Supplier programmes, based on triangulation of perceptions of DFID and Supplier staff. Supplier tender documents were also reviewed, but in many cases this did not give a full picture of the interventions implemented.	4.5

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?	**Primary** • Supplier country case studies **Secondary** • Quarterly verification reports • Supplier monitoring data	Quantitative assessment of % output targets achieved in each country by 2016 (disaggregated by implementing partner if possible). Assessment of the 'scale' targeted by each partner, based on interviews with Supplier staff. Supplier tender documents were also reviewed, but in many cases this did not give enough detail on the scale of the intervention.	6.1.1

DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?	**Primary** • Lead Supplier interviews • LV interviews • Supplier country case studies • Country manager survey • National KII • Local partner interviews • Community-level FGDs and KIIs **Secondary** • Quarterly verification reports	Review of the risks factors cited in the quarterly Verification Reports as affecting implementation in under- and over-performing countries. Preliminary assessment of the main factors affecting implementation across under- and over-performing countries, based on triangulation of perceptions of DFID, Suppliers, and Verifiers.	6.1
DEQ 2.5: Under which circumstances did the PBR framework help/hinder the achievement of intended outputs and outcomes?		Review of the PBR effects cited in the quarterly Verification Reports.	6.1
DEQ 2.6: Under which circumstances did the PBR framework affect the quality of programme implementation (positive or negative)?		Preliminary assessment of the extent to which PBR affected the achievement and quality of outputs (explanation of causal links / external drivers based on staff experience with previous non-PBR programmes), based on triangulation of views across Supplier staff in country and at HQ.	6.1, 6.2
DEQ 2.7: Under which circumstances did Suppliers implement innovative approaches and focus on learning?		Evidence of examples of innovation (compared to staff experience with previous non-PBR programmes), based on triangulation of views across Supplier staff in country and at HQ.	6.1.4
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?	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews	Evidence of transaction costs, based on triangulation of perceptions of DFID, Suppliers, and unsuccessful bidders. Evidence of consequences resulting from transaction costs, based on triangulation of views across Supplier staff in country and at HQ.	2.2.3
DEQ 3.2: Under which circumstances did the PBR modality affect the cost-effectiveness of individual sub-programmes?		Evidence of examples of how PBR affected costs and how PBR strengthened monitoring and evaluation (M&E) arrangements (compared to staff experience with previous	6.4

DEQ 3.3: Under which circumstances did the PBR modality strengthen the programme monitoring and management arrangements of individual sub-programmes?	**Primary** • Lead Supplier interviews • Supplier country case studies • Country manager survey	non-PBR programmes), based on triangulation of views across Supplier staff in country and at HQ. Evidence of examples of how strengthened M&E arrangements led to improved implementation.	6.2.2
DEQ 3.4: Under what circumstances did consortium complexity affect the efficiency of programme management arrangements of individual sub-programmes?	**Primary** • DFID interviews • Lead Supplier interviews • Supplier country case studies	Evidence of how consortium management arrangements affected implementation, based on triangulation of views across Supplier staff in country and at HQ, and DFID.	6.3
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?	**Primary** • Lead Supplier interviews • LV interviews • Supplier country case studies • Country manager survey • National KII • Local partner interviews • Community-level FGDs and KIIs **Secondary** • Quarterly verification reports • WASH sector review for case studies	Preliminary analysis only, with the following planned for 2018: *Health:* Assessment against the prerequisites for WASH health impacts (as set out in Design Document). *Non-health:* Evidence of non-health impacts (explanation of causal links / external drivers).	7.4.3
DEQ 4.2: Under which circumstances did WASH Results Programme activities have any unintended/ unplanned positive or negative impacts?		Only preliminary analysis, with the following planned for 2018: *Review of any unintended effects cited in the quarterly Verification Reports.* *Evidence of factors which caused unintended impacts (explanation of causal links / external drivers).*	6.1.8
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?	**Primary** • Lead Supplier interviews • LV interviews • Supplier country case studies • Country manager survey • Local partner interviews • Community-level FGDs and KIIs **Secondary** • Quarterly verification reports	Only preliminary analysis, with the following planned for 2018: *Review of the risks cited in the sustainability red, amber, green (RAG) ratings in each Verification Report.* *Assessment against the prerequisites for sustainable WASH outcomes (as set out in Design Document), based on triangulation of examples given by DFID, Suppliers, national sector stakeholders and beneficiaries.*	7.4.2

4 DFID's overall programme design and the Supplier's design of individual consortia activities

This chapter sets out the key features of the WASH Results Programme, which is currently half way through its four-year implementation period (April 2014 to March 2018). It describes the key design features set by DFID, and then explains how these parameters were interpreted by each of the Suppliers who implemented activities across 12 countries.

This chapter relates particularly to the 'relevance' aspects of the evaluation.

Key findings of this chapter
• Overall programme objectives are clear, but the theory of change insufficiently articulates how programme activities intend to achieve the intended results, and how the PBR modality contributes to this achievement.
• DFID's rationale for applying PBR to this programme is not fully clear, but the evaluation explores whether the benefits typically expected from PBR interventions have materialised so far.
• The programme design features distinct Output and Outcome Phases, with the latter designed to consolidate efforts in support of sustainability. However, a strong focus on output targets and the 2015 deadline may have inadvertently marginalised sustainability initiatives in the early part of the programme.
• DFID hoped to promote innovative private sector partnerships and interventions through this programme. Most programmes are built on established partnerships, though DFID succeeded in stimulating the formation of new consortia for delivery at scale.
• While programme activities targeted areas with inadequate WASH services, implementation strategies lacked an explicit equity focus: only one Supplier routinely provided results data disaggregated by wealth status. DFID cannot ascertain to what extent people in the lowest wealth quintile have benefitted from the WASH Results Programme.
• While all Suppliers aligned activities with national sector priorities, only SSH4A and some SWIFT projects implemented directly through government agencies. Programme activities haven have been largely successful in strengthening coordination at district level, for example by drawing up a District WASH Plan; however, strengthening of country coordination mechanisms at the national level has not been featured during the Output Phase.
• Under this programme, DFID only paid for achievements up to an agreed ceiling, with payment penalties being imposed for under-performance (down-side incentive). The three Supplier programmes, however, differ in terms of PBR contract design and the definition of 'results' that are eligible for payment. These differences have affected how PBR has played out in practice.
• The verification process has varied significantly by Supplier, being tailored to individual results frameworks.

4.1 Overview

The WASH Results Programme is described by DFID as an 'innovative global portfolio fund'. By the time of completion of the overall design the programme's objective was to deliver sustainable access to clean drinking water sources, improved sanitation facilities and hygiene promotion activities to at least 4.5 million unique individuals by December 2015, with scope to add more thereafter.

It was intended that all 4.5 million people would be reached within the April 2014 to December 2015 period. This reflected DFID's commitment to achieve the MDG targets by December 2015.

However, the programme will continue until March 2018, with additional activities aimed at ensuring that improved access to water/sanitation and hygiene-promoting behaviours are *sustained* up to 2018, and hopefully beyond.

The programme design consists of a hierarchy of three nested designs:

- **DFID's overall programme design** was described in its Business Case, terms of reference and procurement documentation (together with *ad hoc* clarifications of these documents). This is the highest and most general level of the design, and it is described and analysed in this chapter. The period covered is from the initial programme conception in 2012 to the issuing of PQQ and ITN documentation in the second half of 2013.

- **Each of the contracted Suppliers proposed their own programme designs** and results frameworks, two of which related to only two or three countries (SWIFT and SAWRP), and one of which (SSH4A) was implemented in nine countries. This is the second level of programme design, described in this chapter, and the period covered is from late 2013 to the completion of contract negotiations in April 2014.

- For each Supplier, the **design of the programme varied by consortium membership, programme components and country.** This represents a third level of detail, which, together with the differences between country level implementing partners, provides the context for the evaluation findings set out in Chapter 6. The period covered is that of the programme implementation, namely April 2014 to March 2018.

Table 6:   Intervention logic of the WASH Results Programme

	Outputs	Outcomes	Impacts
Targets	• 1.2 million people with access to clean drinking water sources • 4.3 million people with access to an improved sanitation facility • 9.3 million people reached by hygiene promotion activities	• 75%–100% of the people reached by December 2015 continue to use clean drinking water sources • 75%–100% of the people reached by December 2015 continue to use improved sanitation facilities • 10%–15% of the people reached by December 2015 use hygiene-promoting behaviours	Poor people benefitting from the WASH Results Programme have improved health status, as measured by reduced diarrhoea and child mortality
Time period	By December 2015	By March 2018	By March 2018*

*Suppliers are not required to report on impact targets and these are not linked to PBR payments.

4.2   Features of DFID's programme design

The most prominent feature of the WASH Results Programme is its funding modality, namely PBR. However, other aspects also distinguish it to a greater or lesser extent from other WASH programmes, and these too are examined in this chapter.

At a high level, DFID's design is summarised by its theory of change (see Figure 2) – the 'what' and 'why' – together with the arguments put forward in the Business Case about the financing modality – the 'how'. DFID's PQQ and ITN documents and clarifications provide further evidence of DFID's design thinking in regard to the programme. A critique of the theory of change is presented in Annex B.

In terms of its structure and shape, the important features of the WASH Results Programme are as follows:

- **Scale**. Each Supplier contract was relatively large, averaging £25 million. This was a similar order of magnitude to the UNICEF ASWA programme, to which DFID provided a budget of £35.5 million.

- **Phasing and timing.** The WASH Results Programme was constructed in two phases: the Output Phase, in which the main outputs were to be delivered between April 2014 and December 2015; and an additional two years (the Outcome Phase) to consolidate efforts in support of sustainability, taking the programme to March 2018. The 'hard deadline' of December 2015 (the close of the MDGs period) marked the transition from Output to Outcome Phase.

- **The PBR financing modality.** Payments to Suppliers are triggered by the independently verified delivery of pre-specified 'results' (key activities, outputs and outcomes). The PBR model gives Suppliers freedom to design and implement their own approaches, so long as the agreed results are delivered. (Details of the PBR modality are provided in Chapter 5).

- **MV.** The MV component is a vital element of the PBR model, as it is responsible for independently verifying the Suppliers' delivery of the contractually specified 'results', which in turn triggers payment from DFID. The evaluation is expected to generate important lesson learning across a range of issues, including for large-scale WASH programming, how best to apply PBR in WASH and applications of PBR more generally.

- **Innovative private sector partnerships.** There was an expectation that 'innovative' private sector partnerships would lead to important learning, especially for NGO partners.

- **Innovative interventions.** There was an expectation that the PBR modality would foster innovative interventions as Suppliers would be free to employ whatever approaches they want in order to achieve their results targets.

- **Counting beneficiaries.** DFID's guidance (methodology notes for WASH) for water supply and sanitation is similar, with both referring to numbers of people gaining access to the service. This contrasts with its guidance on hygiene promotion, in which the numbers of beneficiaries are those reached with hygiene messages (not necessarily those with access to appropriate facilities, or those that use such facilities). Water supply and sanitation improvements in health facilities, schools and other public places are deemed ineligible for payment; only unique individuals receiving services are to be counted.

- **Sustainability**. All guiding documentation for the programme acknowledges the importance of services and behaviour changes standing the test of time. However, some other programme design aspects may not be best suited to achieving sustainability.

- **Equity**. There were mixed messages in the design documentation in regard to equity, resulting in the varying degrees to which this outcome was addressed in the Suppliers' designs and programming.

- **Impact**. The theory of change identified improved health, specifically in relation to diarrhoea and under-five mortality, as the ultimate goal of the programme.

In the following sections and sub-sections, each of these programme features is addressed in turn. First, however, some key aspects of the overarching programme concept are examined.

4.3   Overarching programme targets

4.3.1   Targets for WASH outputs

Overall output and outcome targets were specified by DFID, with each Supplier allowed to bid according to one of two 'lots':

- two contracts were awarded under **Lot A (water, sanitation and hygiene)**: aiming to reach at least one million people reached per contract; and

- one contract was awarded under **Lot B (sanitation and hygiene),** aiming to reach at least 2.5 million people reached in total (of which 1.75 million people to be sanitation beneficiaries).

While these targets seem straightforward at first glance, Lot A needed some clarification in terms of how one million beneficiaries would be calculated, given that the programme included three components and there would, almost certainly, be some overlap between those gaining access to sanitation and those receiving hygiene promotion and/or gaining access to water. DFID's protocols for counting beneficiaries emphasise the importance of counting unique individuals:

> *… the WASH Results Programme aims to deliver improved WASH services to 4.5 million unique individuals by December 2015. Each person may be reached with one or more interventions in water, sanitation and/or hygiene but each individual must only be counted once, even if reached with more than one water, sanitation and/or hygiene intervention.* Terms of Reference, clause 1.11.

While this basis for counting beneficiaries was explicit in the DFID guidelines, in practice it has been difficult for the MV team to verify the number of unique beneficiaries in those instances where unique household ID numbers are not available (these were only available for SAWRP). Some double-counting risks remain for SWIFT, although this has been recognised as an issue that SWIFT are addressing (see box below).

The total number of unique beneficiaries reached was counted based on the reach of water and sanitation interventions. While DFID pays for some stand-alone hygiene promotion interventions in this programme, these are not counted as results against DFID global WASH targets. It is also notable that services provided in schools, clinics and other public settings did not count towards DFID's global WASH targets target. These requirements may need to be revised for future programmes, given especially that the Sustainable Development Goals include targets for WASH services in schools and health facilities.

> **Box 2: The challenge of avoiding double-counting**
>
> There are two distinct double-counting risks under the WASH Results Programme:
>
> i. reporting the same facilities as new in successive reporting rounds, so that the associated beneficiaries are counted multiple times, and
>
> ii. over-stating the number of unique beneficiaries if the WASH beneficiaries are simply added, when the same person may receive more than one type of benefit (i.e. water, sanitation and/or hygiene).
>
> The verification process placed emphasis on preventing the former by working with Suppliers to ensure that they had procedures to prevent such instances. For SAWRP, it was ensured that a unique ID number was assigned to every household with a new latrine. Spot checks were carried out to ensure these procedures were in place. For SSH4A, spot checks were carried out to check whether the same person was named twice in the attendance lists of hygiene behaviour change events. For SWIFT, the nature of the contract (with several distinct consortium members, each with their own approach) has made it more difficult to check if the same person was counted multiple times, but this is an issue that SWIFT are addressing.
>
> With regards the latter, some SAWRP and SWIFT interventions were specifically designed to deliver hygiene and sanitation and water to the same person. The risk of double-counting in such cases appears minimal under SAWRP as unique beneficiary ID numbers can be used to disaggregate results where required. This risk is higher for SWIFT, where unique ID are not available, although this has been recognised as an issue that SWIFT are addressing. For SSH4A the risk is also minimal as water provision was not part of their intervention, and all sanitation beneficiaries can be assumed to be a sub-set of the substantially larger group of hygiene beneficiaries.

4.3.2   DFID's programme ambitions at outcome level

Outcome targets were not specified in the request for proposals: DFID stated that the use of WASH facilities, and the adoption of hygienic behaviour, would be sustained *among* the total number of beneficiaries reached at output stage. Suppliers presented their level of ambition for the transition from outputs to outcomes in their results frameworks, but these targets were somewhat provisional and were not finalised until late in the Output Phase, following communication from DFID on what would be a minimum acceptable level of achievement. A key challenge for Suppliers was that there were no established benchmarks in the WASH sector for the conversion of outputs into outcomes (for example, the percentage of those receiving access to water or sanitation services who may realistically be expected to still enjoy use of that service at some subsequent date). Based on experience, Suppliers could make rough projections regarding likely levels of water-point functionality and use two years after installation, but this was much harder in regard to the functionality and use of toilets, and the conversion of those who had received hygiene messages into those actually practising hand-washing with soap.

4.3.3   The level of ambition chosen by Suppliers

At the bidding stage, DFID provided a long list of countries for Suppliers to choose from but did not specify how high targets should be set in each country. Each consortium decided at the tender stage what level of WASH outputs could be realistically achieved by December 2015, and (provisionally at least) what level of outcomes could be sustained until March 2018.

In bidding for the work, each Supplier had to commit to meeting DFID's minimum output targets overall. For each participating country, Suppliers assessed what was doable based on: how quickly they could scale up their existing operations; current levels of WASH access; the implementation capacity of local partners and/or local government; and the potential for sustainable outcomes.

Each consortium was careful to commit only to targets that they were confident of meeting. Three elements played a key role in determining the ambition of Supplier targets:

- The fact that the **PBR modality** did not reward over-achievement but penalised under-achievement led Suppliers to propose *conservative* targets, given the short time available for the Output Phase, and given that payments were linked to the degree to which results could be sustained until programme end (see Section 6.1.1)

- **Ambitions were matched to the type of implementing partner chosen:** The SSH4A programme implements directly through governments in several countries, and was therefore very conservative in setting targets, to match the implementation capacity of local government. In contrast, Unilever implements exclusively through local partners and set their ambition based on their own organisational capacity and risk appetite, which was very high in Bangladesh. Unilever explained that, with hindsight, they could have set higher targets for Pakistan. Another consortium member commented that they encouraged local partners not to view existing staffing levels as a limiting factor but to build additional staffing into their price.

- **The baseline level of WASH access in country** was another important factor in setting targets. SNV explained that while initial ambitions were based on rough baseline data from 2012, these ambitions were revised in some countries once an updated baseline survey was undertaken at the start of this programme in June/July 2014. Similarly, SAWRP significantly revised ambitions once the baseline survey in Pakistan revealed significantly higher baseline levels than partners had estimated. As a result, local partners had to increase their geographical scope: this required larger field teams, more transport, and other support. As a consequence, local partners were obliged to renegotiate financial aspects of their agreement with their consortium partner, as PBR implementation budgets were held at the country level, not globally.

Given that all almost all output targets were eventually met (see Section 6.1.1), it is evident that that they were set at a realistically achievable level. It remains to be seen, however, whether outcome targets are realistic and whether DFID has allowed the bar to be set too high or too low.

4.4   Theory of change – The 'what' and 'why'

DFID's theory of change diagram for the WASH Results Programme, as included in the Business Case, is reproduced as Figure 2 below. It is fairly rudimentary and in essence is a logframe on its side. The content is reviewed briefly below (and more fully in the Evaluation Design Document). Suppliers were in fact required to develop their own theories of change, but to our knowledge did not do so, primarily because the focus of each design was the Supplier-specific Results Framework against which payments would be made. The results frameworks have a high profile in programme operations and in practice the theory of change has not, to our knowledge, been a key point of reference either for DFID or the Suppliers. Discussions around the adequacy of the programme theory of change are therefore somewhat academic; nevertheless, in Annex B1 we offer some analysis on the adequacy of the theory of change since it underpins programme design. In brief, we note that the narrative component of the theory of change lacks depth and essential detail, especially on programme context; what or who needs to change, and why; how this change will be brought about; and what key assumptions underlie programme design. Moreover, it makes no reference to the PBR modality, how it will be applied and what should happen as a result of its use. Notwithstanding, it is noted that the page limit of the Business Case likely constrained the level of explanatory detail that could be included on the theory of change.

In Annex B.2, we examine the generic theory of change underlying PBR set out in DFID guidance notes. This has been a point of reference for the evaluation as it identifies the effects that should be expected from the adoption of a PBR modality, though these are not WASH-specific.

Figure 2:   DFID WASH Results Programme theory of change
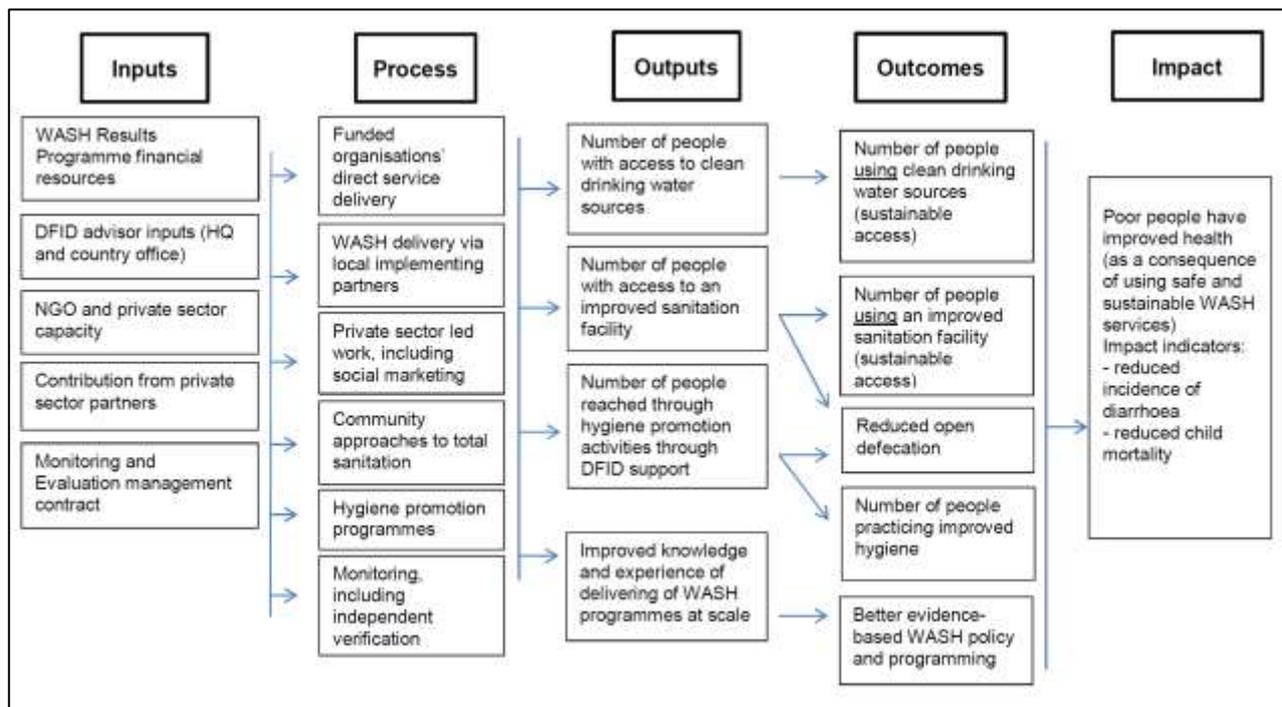
4.5   Brief summary of each Supplier

4.5.1   SAWRP

Implementing organisations

The SAWRP consortium is led by Plan International, with WaterAid, Unilever and WSUP as members; local NGOs as implementing partners; and WEDC and Ipsos as advisory partners. Activities are implemented in **Bangladesh** and **Pakistan,** each of which has a Consortium Coordination Unit with a Programme Director, MVE Specialist and Finance Coordinator.

Country context and targeting of activities

SAWRP projects are aligned with Pakistan's national sanitation strategy, known as the Pakistan Approaches to Total Sanitation (PATS) and Bangladesh's national WASH strategy. National quality standards and community-led total sanitation (CLTS) procedures are being followed. In Pakistan, the government has not been directly involved in implementation: it is involved in ODF certification and will be involved in sustainability monitoring in the Outcome Phase. The programme is also supporting government-led WASH coordination at local government level.

Baseline access to rural WASH services and hygiene knowledge levels were relatively high in both countries, but this still left a huge number of people under-served, given the high populations. The high population density in both countries enables SAWRP to reach a very large number of

beneficiaries. SAWRP targets harder-to-reach communities that have not been covered by previous WASH interventions, for example flood-prone areas and schools in poorer areas.[5]

WASH activities

SAWRP comprises two independent workstreams:

- Plan International, WaterAid and their local implementing partners promote sanitation and hand-washing with soap in poor rural communities using CLTS and sanitation marketing. They aim to deliver results across entire sub-districts, with an Output Phase target of 227,750 water, 1,644,176 sanitation, and 6,609,570 hygiene beneficiaries by December 2015. ODF is an explicit aim, and hygiene promotion (hand-washing with soap) targets the same communities through a mixture of household visits and mass meetings. Water supply improvements are not a major feature of the programme (only 3% of beneficiaries) but were included in order to provide access to some of the poorest and most under-served poor communities, where communities face acute water supply shortages. WaterAid considered that a substantial water supply component was not feasible under PBR due to the relatively high unit cost of hardware investments. Plan International and WaterAid generally implement in different districts.

- Unilever deliver a hand-washing campaign in a large number of schools in both countries, though the focus is overwhelmingly in Bangladesh, which accounts for 84% of Unilever's beneficiaries. This component has a profound effect on the results profile overall, as it delivers 75% of total hygiene beneficiaries – all of them school children. This said, DFID does not count hygiene-only results when reporting against its global targets.

Figure 3:   WASH targets for SAWRP countries
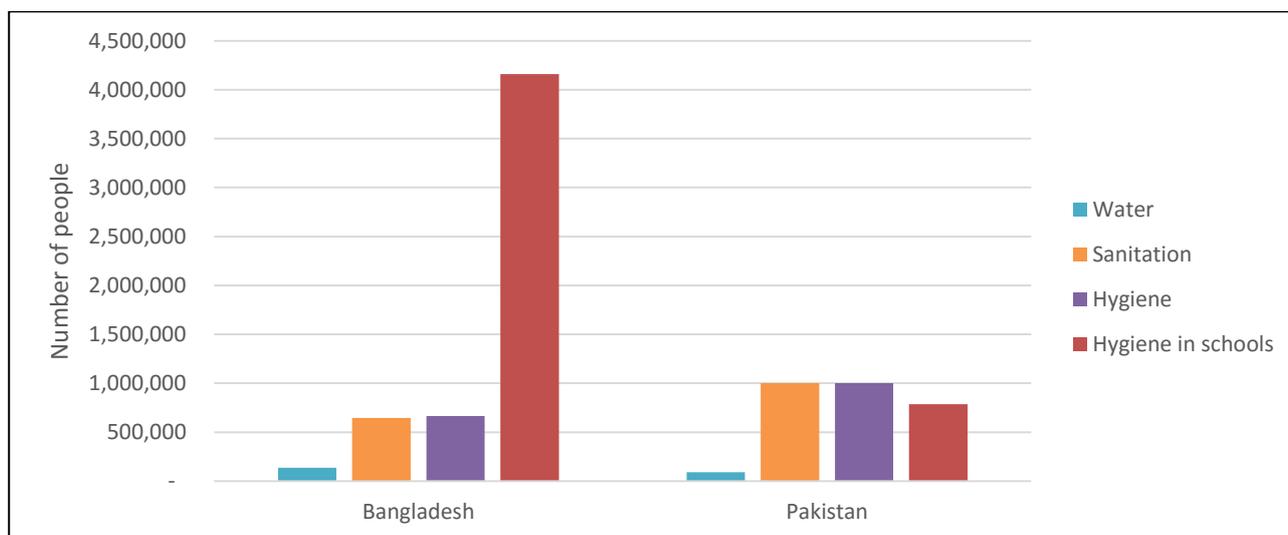
All partners are scaling up recent or ongoing WASH initiatives. In Pakistan, for example, Plan International and WaterAid are building on their earlier work under the UNICEF-led Rural Sanitation in Flood-Affected Districts (RuSFAD) programme, which ran from 2010 to 2012, while Unilever is expanding its ongoing '*School of 5*' hand-washing campaign in both Pakistan and Bangladesh. All partners have applied tested operational models and local partners have some experience of working at scale as part of flood relief projects.

While most of the programme is implemented via local NGOs, Plan International in Bangladesh opted to work directly with local government staff.

---

[5] Unilever report using geographical targeting to identify poorer neighbourhoods. The evaluation team were not able to establish the exact poverty criteria used.

Consortium members have established relationships with the local implementing partners, who in turn have strong links with the targeted communities. While the lead organisations are subject to the PBR modality and have to pre-finance their work, local implementing partners are protected from PBR risks and are funded through regular grant-based agreements.

Monitoring of results and the sustainability of results

Results are monitored though paper-based and mobile-based systems in Pakistan and Bangladesh, respectively. Local partners collect output-related data quarterly, which is collated into the mPIMS system, which provides each water-point and latrine with a unique ID. At the end of the Output Phase these data were consolidated into an Access database in Bangladesh and an Excel database in Pakistan.

In the Outcome Phase, the sustainability of water-points and latrines will be monitored via local NGOs' regular systems; outcome achievement will be assessed via commissioned mid-line and endline surveys; and learning and review will be facilitated via a Sustainability Assessment Framework developed by WEDC.

4.5.2   SWIFT

Implementing organisations

The SWIFT consortium is led by Oxfam GB, the other members being Tearfund and the ODI, which leads on MEAL. At country level the implementing partners are INGOs and include Concern Worldwide, Sanergy, WSUP, BBC Media Action, Practical Action and Africa Ahead. SWIFT operates in **DRC and Kenya;** Liberia was initially an intervention area but was dropped in 2014 due to constraints related to the Ebola crisis. This consortium has the largest number of implementing partners and is managed through national consortium managers sitting within Oxfam Kenya and Oxfam-DRC.

Country context and targeting of activities

In DRC, SWIFT is implemented within the framework of DRC's national sanitation programme, *Villages et Écoles Assainis* (Healthy Villages and Schools, or VEA). While government staff are not involved in implementation, they are involved in ODF certification and will be involved in sustainability monitoring in the Outcome Phase. SWIFT in Kenya is not linked to any specific government programme or strategy, though it has encouraged recently created county governments to play an active role.

Baseline access in Kenya was mixed: in the arid and semi-arid lands (ASAL) region of northern Kenya the level of access to WASH facilities is very low. In small towns and urban Nairobi, the level of WASH access is higher but partners have focused on chronically under-served slums and informal settlements.

While there are varying definitions of 'fragile' states and contexts, the term really only applies to SWIFT's work in DRC and the ASAL region of Kenya, both of which have ongoing security concerns. The context for the urban work in Kenya is acute poverty, weak public institutions, and constraints relating to land tenure and access to public services, but not environmental fragility or conflict.

WASH activities

SWIFT is the only consortium contracted under Lot A, meaning that overall the programme has a substantial water supply component (one-third of beneficiaries), in addition to sanitation and hygiene promotion (two-thirds of beneficiaries). In contrast to the other Supplier programmes, activities encompass both urban and rural contexts. In Kenya, the activities include:

- CLTS in rural areas, including the provision of free slabs in one Kenya project (Oxfam in Turkana);
- rural hygiene promotion, within and beyond the communities targeted with CLTS;
- rural water supply improvements, including drilling and rehabilitation of boreholes equipped with solar pumps (Oxfam, Practical Action, Concern);
- urban water supply in Nairobi slums (WSUP, Practical Action);
- urban sanitation promotion in Nairobi slums using a social enterprise model, plus school hygiene promotion (Sanergy); and
- hygiene promotion radio broadcasts through mentoring small private and community-based radio stations (BBC Media Action).

SWIFT in Kenya is not really a single programme operating at scale, but a portfolio of small- to medium-sized projects, which, taken as a whole, contribute a significant number of results towards overall programme targets. Some of the projects are stand-alone initiatives unrelated to others in the consortium, and rural sanitation targets are very modest compared to DRC and other countries in the global WASH Results Programme. Local NGO partners have a limited role, being involved in selected rural sanitation and hygiene promotion projects led by Oxfam and Concern, and are not on PBR contracts. In the case of urban water supply, SWIFT partners work closely with utilities. An unusual addition to the programme is BBC Media Action support to community-based radio stations, which make hygiene promotion broadcasts.

The programme in DRC is quite different – here a smaller number of partners deliver results on a much larger scale.

Most local partners have worked in their intervention areas before and have strong links with communities.

Figure 4:   WASH targets for SWIFT countries
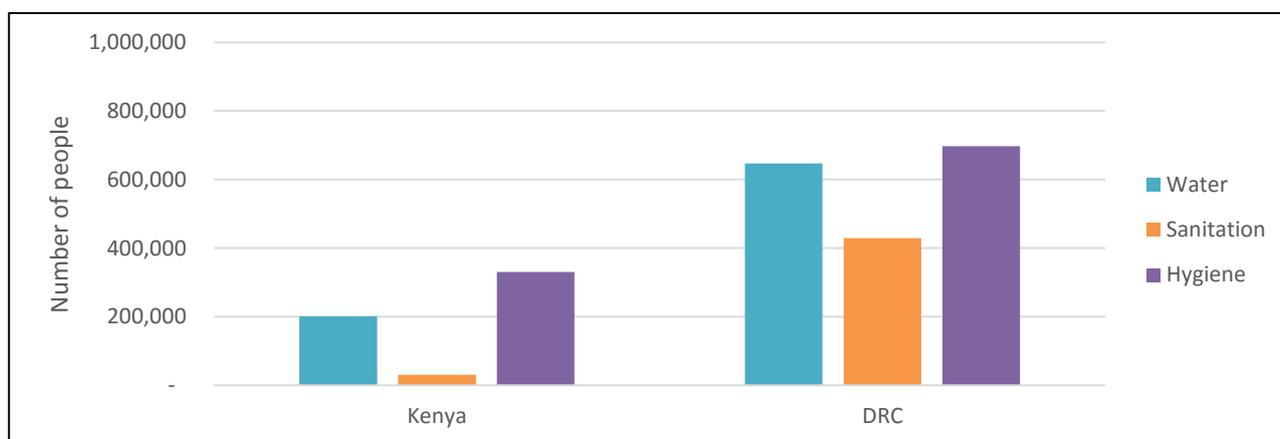
Apart from the number of partners and range of intervention types, another key difference between SWIFT and the other two programmes is that not all the operational models were tried and tested at the programme start, or ready for scaling up. During the Output Phase the rural sanitation programme in Kenya did not gain much traction.

Monitoring of results and the sustainability of results

The partners in Kenya programme had existing, paper-based monitoring systems at the start, and most of these needed some improvement to meet MVE requirements. In DRC, a bespoke monitoring system was developed at the outset.

The use and sustainability of WASH facilities will be monitored through household surveys in the Outcome Phase.

4.5.3   SSH4A

Implementing organisations

SSH4A is solely implemented by SNV and its local partners – there are no other consortium members. Activities are being implemented in **Ethiopia, Ghana, Kenya, Mozambique, Nepal, South Sudan, Tanzania, Uganda and Zambia**. SSH4A is managed from a programme management unit in Nairobi.

Country context and targeting of activities

As government alignment is a key element of SSH4A, activities in all nine countries are well-aligned with national WASH frameworks and implemented through government staff where possible. The exact nature of activities is tailored to each national WASH context: in 'early adopter' countries, where progress towards ODF has been strong, SNV has focused on the remaining harder-to-reach communities (Kenya, Nepal, Uganda and Zambia). In Tanzania, SNV has focused on the upgrading of sanitation facilities and on hygiene promotion, as the country has low levels of open defecation but high levels of unimproved facilities with sustainability and hygiene challenges. In Ghana, SNV has focused on targeted behaviour change communication (BCC) to reduce the number of shared latrines. In 'late adopter' countries like Mozambique and South Sudan, SNV has focused on providing specific support to build the basic capacity of the CLTS movement. The country contexts are generally stable, though security issues posed challenges for implementation in South Sudan, Nepal and Mozambique.

Some countries had high baseline levels of WASH access, whereas others had very low levels of access when programme activities were being designed in 2013. National-level Joint Monitoring Programme (JMP) data (see below) give an overview of the variation in WASH access at programme start, with the caveat that JMP data are nationally representative and that programme activities may have been planned in certain regions which deviated significantly from national averages.

Figure 5:   JMP estimates of sanitation coverage for the nine SSH4A countries
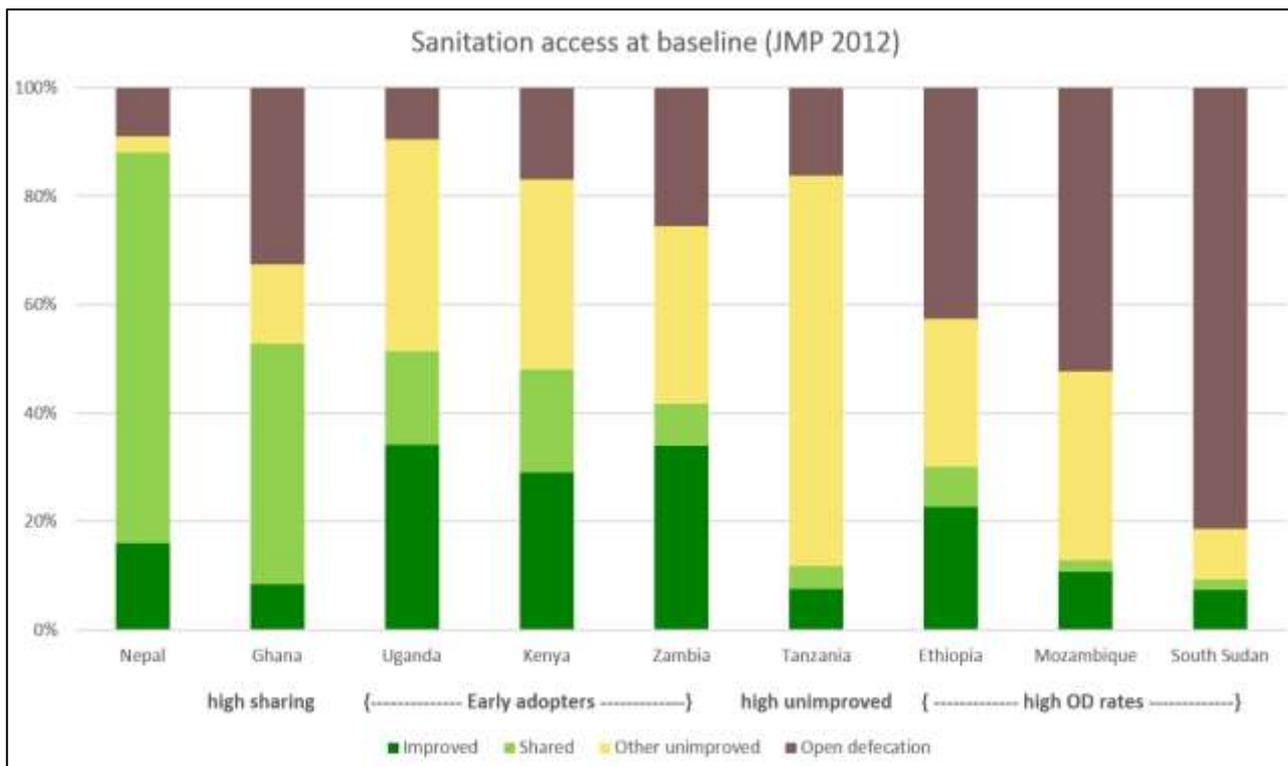
Note: the classification of 'early adopters' was added by SNV; it does not feature in the JMP data.

WASH activities

This programme has no water supply component – it deals only with rural sanitation and hygiene promotion. Activities are implemented within the framework of the existing SSH4A programme, which operates in a dozen countries in Asia. It embodies a tested operational approach for working at scale and combines demand creation − mostly using CLTS − with support to sanitation supply chains, BCC and strengthening capacity for WASH governance. Attainment of ODF is a key objective.

Figure 6 shows the targets which SNV committed to reaching in each country. In practice, each country programme aimed for a somewhat higher target, providing SNV with a safety margin should delivery of results prove to be more challenging than was anticipated.

Figure 6:   Sanitation and hygiene targets for SSH4A countries
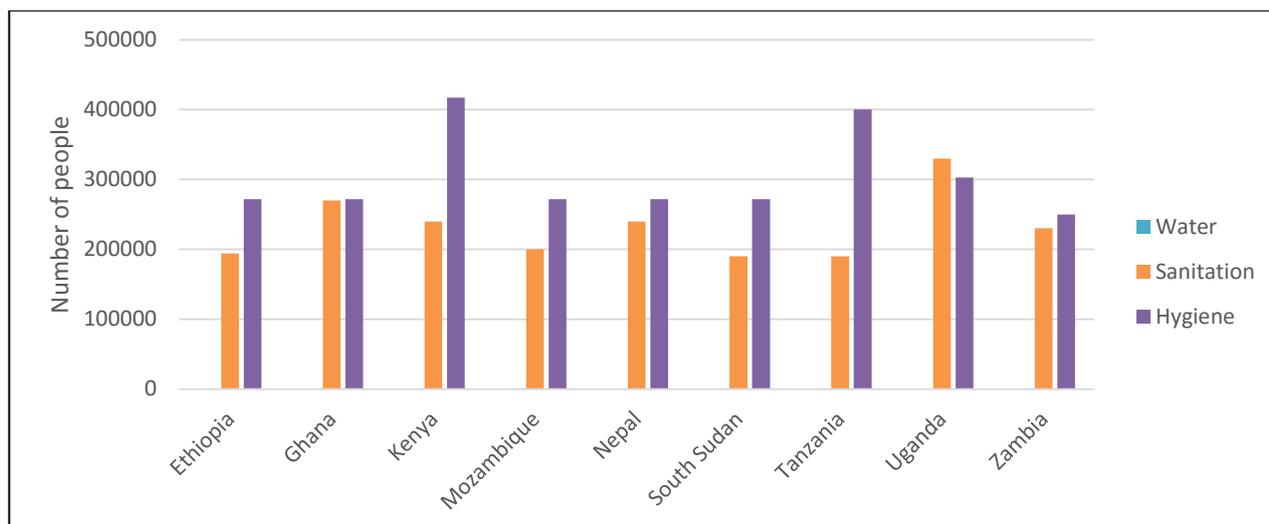
In several countries, activities are implemented by local government partners, supported to a greater or lesser extent by NGOs referred to as Local Capacity Builders (LCBs). These include a mix of existing and new SNV partners, and some (including those in Uganda) are contracted on a PBR basis.

Monitoring of results and the sustainability of results

Results are being monitored through household surveys, administered through the tried and tested Akvo Flow mobile-based monitoring system[6].

During the Outcome Phase, sustainability aspects will be monitored through 10 indicators which track the existence of various technical, environmental and institutional prerequisites. Some of these indicators are being monitored already.

4.6   Programme scale and timing

4.6.1   Scale of operations

The WASH Results Programme was conceived of as a large-scale overall intervention, with 'at scale' interventions by individual Suppliers. Our understanding is that 'at scale' in this case simply referred to the total beneficiary numbers, not to the operational strategy. DFID did not, for example, specify that activities had to cover entire districts or sub-districts, or that programmes should operate via the government framework.

4.5 million people were to receive first-time access to either water supply, adequate sanitation and hygiene messages or to sanitation and hygiene only, by December 2015. In terms of the global need (approximately 700 million people without adequate water services and 2.5 billion without sanitation at the time of design in 2012), the programme was envisaged as making a small contribution. Nevertheless, each Supplier programme was much larger than typical INGO WASH programmes. DFID have indicated that stimulating the establishment of NGO and private sector partnerships for delivery at scale was in fact the key innovation sought via the programme.

Most Suppliers expanded, or built upon, earlier WASH programmes, and could therefore utilise tested operational approaches and established institutional relationships with local government

---

[6] See http://akvo.org/products/akvoflow/.

agencies and NGOs/community-based organisations. Adopting tested models and selecting a limited range of projects and partners enabled Suppliers to achieve economies of scale. SAWRP's design clearly aimed for delivery at scale as it aimed to achieve ODF status across the entirety of many Union Councils in Pakistan, and worked via government in Bangladesh (at least for part of the programme). Similarly, SSH4A aimed for district-wide coverage in six of its nine countries, and this entailed a significant role for local government.[7]

SWIFT also aimed for district-wide coverage in DRC, but in Kenya rural sanitation activities were on a fairly small scale. SWIFT in Kenya is in fact a package of small-to medium-sized projects that operate largely autonomously, though Oxfam did much to ensure that each partner was on track during the Output Phase, and that there was good communication between them. The range of project types in Kenya mean that the programme did not offer economies of scale, and the transaction costs of management and coordination have been relatively high. SWIFT in Kenya was also a rare case among the country programmes in that some partners used operational approaches that were not fully tested and ready for scaling up at the programme start. Despite these challenges, the country programme succeeded in meeting its Output Phase targets.

4.6.2   Phasing and timing in the global design

Two comments are made here in regard to the phasing and timing of the programme.

First, in terms of overall programme design, the MDGs deadline of December 2015 was the determining feature of the programme structure. In order to contribute to the UK government's MDG commitments, the numerical beneficiary targets had to be achieved by that date. The post-2015 period can be interpreted in at least two ways. During the public consultations in the design phase the main concern voiced by some WASH professionals was that adding a programme phase to consolidate efforts in support of sustainability, only *after* first-time services had been provided in the Output Phase, was counter to good practice: sustainability should be addressed from the outset in order to have a realistic chance of being realised.

The alternative point of view, articulated by DFID, was that the inclusion of a post-implementation phase was an innovative and positive design feature. DFID expected that measures in support of sustainability should feature in plans and operations from the start of each project. Additionally, having a dedicated Outcome Phase would mean that extra time and resources were available to consolidate sustainability measures introduced during the Output Phase. DFID's design did not intend to marginalise sustainability measures in the early part of the programme. In hindsight, for some local partners, the strong focus on output targets and the 2015 deadline inadvertently marginalised sustainability initiatives in the early part of the programme – as explained in more detail in Sections 6.1.5 and 6.1.6.

Secondly, it is important to note, however, that as the programme design and procurement period extended, so the implementation period steadily shrank. The overall programme duration was reduced from five to four years as a result of the time it took to complete the design and procurement processes. This had implications for the quality and likely sustainability of services provided, as we highlight further in Sections 6.1.5 and 6.1.6.

---

[7] The exceptions were Uganda, Kenya and South Sudan.

4.7 Financing modality – The 'how'

4.7.1 Justification for the PBR modality at design stage

DFID clearly sets out in its technical guidance on PBR that '*if designed and delivered well, potential benefits of PBR include: Improved efficiency and effectiveness of delivery from the sharpened incentives in a PBR contract (paying once results are delivered rather than upfront); Increased potential for innovation and flexibility in delivery through not specifying how to achieve results; Increased transparency and accountability for results; and a stronger focus on performance*' (DFID 2014b).

However, when the Business Case for the WASH Results Programme was written in 2013, it was not clear how these potential benefits would apply to the WASH Results Programme. The Business Case listed 10 possible options for the financing modality of the WASH Results Programme, and analysed five (including 'do nothing') in greater depth. However, the full reasoning and assumptions underlying the final selection of a PBR financing modality (and the particular form of the modality employed) were not made clear. In Annex B2 we set out in diagrammatic form our own construction of DFID's generic reasoning for using PBR, and a lengthy list of assumptions – some or all of which presumably (for this is not made explicit) underpinned the programme design. These reconstructions raised a number of questions:

1. What exactly was the implementation or programming problem which a PBR financing modality was meant to address?
2. At the design stage, were assumptions made about the types of organisations which might bid to participate in the programme: for example, the assumption that private sector organisations, presumed to need financial incentives to perform well, would bid—possibly successfully?
3. Was the selection of a PBR modality as much to do with DFID wanting robust data regarding results as to do with enhancing programme effectiveness and efficiency?
4. Was there an assumption that generally accepted per capita costs in grant-financed programmes were too high, and that PBR would drive better VFM?
5. Or was the selection of the PBR modality as much a public policy and political choice as other aspects of the programme, such as the 60 million WASH beneficiary commitment and the MDGs deadline?

Some of these questions are explored further in Chapter 8. We also acknowledge that DFID has recently clarified that the primary reasons for using PBR in this case were a) to stimulate the creation of new organisational partnerships for delivery at scale and b) to achieve a high degree of certainty regarding the number of beneficiaries that would be reached, and when.

4.7.2 Variations in how 'results' were defined by Suppliers

Each Supplier set out which results they would be paid for, and when, in a 'results framework'. DFID did not specify which results could and could not be paid for under a PBR contract, only indicating that '*the payment mechanism will need to be intrinsically linked to the delivery of the output results and the sustainability aspects of the Programme*'.[8]

In the event, DFID were quite flexible in negotiations and agreed to some payments for processes (inputs) as well as outputs (see Table 7 and Figure 7). These negotiations, however, led to variations in contract specifications. Payments for inputs and intermediate outputs are less common in other PBR programmes and DFID's PBR guidelines recommend that indicators for

---

[8] Supplier Terms of Reference, October 2013.

payment should be as close to outcome level as feasible (DFID 2014b). However, in this programme Suppliers saw payments for inputs and intermediate outcomes as essential in order to secure some cash flow in the earlier parts of the programme, given that this was a 100% PBR programme. We acknowledge that DFID's policy on PBR in WASH is still evolving and has not set limits on the types of results which payments can be claimed against.

SAWRP had the most 'pure' PBR contract, with results defined only in terms of beneficiaries reached, while SWIFT and SSH4A also included a series of workplan activities, such as the signing of water-point construction contracts, development of BCC materials and the training of facilitators (referred to as 'process' results hereafter). SSH4A additionally included the sign-off of district sanitation plans with local government, which can be considered a prerequisite for achieving sustainability. In addition, both SWIFT and SSH4A listed their management fee and/or the fee for their M&E contractor separately.

When the programme began, it was made clear that no additional payments would be made where Suppliers over-delivered against their contracted targets (upside incentive), though there would be penalties for under-performance (down-side incentive). Interestingly, DFID later offered funding for results beyond the original targets in some instances, but Suppliers declined. It also appears that partners, like WaterAid Bangladesh, paid incentives for over-achievement to their in-country implementing partners.

Table 7:   Comparison of contracted 'results' across Suppliers

Result type	SAWRP	SWIFT	SSH4A
Input		M&E contractor (ODI) Management fee	Management fee
Process		Water-point construction contracts signed; water management structures set up Sanitation promoters trained; number of communities triggered BCC materials completed; hygiene promoters trained	Baselines completed Workplans agreed District sanitation plans in place BCC materials designed Training of facilitators
Output	Number of people *with access to* W/S/H	Number of people *with access to* W/S/H	Number of people *with access to* S/H
Outcome	Number of people *using* W/S facilities or practising hygiene behaviour	Number of people *using* W/S facilities or practising hygiene behaviour	Number of people *using* sanitation facilities or practising hygiene behaviour

Figure 7:   Comparison of payments made towards different results across Suppliers

Note: 'Inputs' here refers to direct payments to international partners supporting monitoring, knowledge management and learning but not directly involved in programme implementation. 'Process' refers to intermediate steps towards the delivery of outputs, such as community mobilisation, training and the development of IEC materials.

The variations between how 'results' were defined under different contracts led to substantial discussions between DFID, the MV team and the Suppliers on verification procedures and on how to define the hygiene output result[9], particularly with regards to the number of people 'reached' by hygiene promotion. In the sector, there is no widely accepted definition of what 'reached' means and what a minimum level of performance looks like. The MV team issued a number of notes to help clarify good practice in other aspects of hygiene promotion, but it was not possible to provide a single definition of 'reached' that was applicable across the wide range of hygiene-related activities undertaken by the three Supplier programmes. As a result, each implementing partner agreed separate definitions with the MV team and this entailed lengthy debate:

- For example, in the case of SWIFT there were protracted negotiations on how to count beneficiaries of hygiene-related radio broadcasts when these were made in the same location as that where other partners were delivering hygiene promotion on the ground.
- In regard to being 'reached' by interpersonal hygiene promotion, SWIFT defined this as, as a minimum, that a person received a household visit and attended a group / mass meeting.
- In SAWRP, provided there was evidence that all planned hygiene promotion activities had taken place in a community, one hygiene beneficiary was counted for each person gaining access to sanitation.
- SNV, meanwhile, relied more on surveys testing respondents' recall of hygiene messages. There is a risk that such surveys captured prior knowledge, without adequate differentiation in respect of recent SNV interventions.

---

[9] This highlights a potential limitation of the OJEU negotiated tendering process.

4.8   Enabling innovation, equity, sustainability and impact

4.8.1   DFID's emphasis on innovation

DFID's Business Case includes expectations of the emergence of (a) innovative private sector partnerships, and (b) innovative WASH interventions. There was an expectation that innovative private sector partnerships would lead to important learning, especially for NGO partners, and that the PBR modality would foster innovative interventions as Suppliers would be free to employ whatever approaches they wished in order to achieve their results targets.

> *An early decision was to combine the roles of the private sector with those of NGOs, and not to treat them separately. There were long discussions in DFID about what the UK private sector could offer. The general feeling was that the WASH expertise lay within NGOs and academia but that the private sector could bring useful financial and overall management capability. Not many WASH NGOs can deliver large-scale programmes in a short timeframe. Consequently, a combination of private sector management and NGO expertise could be strong. In principle, also, the private sector could bring its financial assets into play in results-based financing, when NGOs don't have these reserves to draw on. – DFID WASH adviser*

There is potential confusion between **innovation** as something new or different, and **flexibility** within an overall guiding framework. DFID provided the space for Suppliers to propose approaches and methods of their choice which would meet DFID's beneficiary targets and fulfil the other criteria specified in the programme terms of reference. There was therefore flexibility for Suppliers at the design stage, although once the results frameworks were agreed and contracts signed that flexibility would inevitably be reduced.

Innovation in the operational approaches adopted for programme implementation was less likely to occur. Suppliers have been well aware that they were taking on unusual and demanding financial risks, resulting in a reluctance to implement new and untested implementation approaches. In practice, operational innovation has not featured much in the three Suppliers' programmes, though there have been some examples of adaptive learning during implementation, as well as innovation around MV (see Section 6.1.4).

DFID have recently clarified that the content of the Business Case was somewhat provisional and much changed in the course of Supplier procurement, not least because DFID adopted a negotiated approach and was flexible regarding some aspects of the detailed design. DFID's stance on innovation has also been clarified, indicating that stimulating the formation of new organisational partnerships for implementation at scale was also a priority for the programme. These partnerships were indeed formed, though private sector participation has been limited, with only SWIFT having a private sector partner (Unilever).

4.8.2   DFID's emphasis on sustainability

DFID's programme design placed a strong emphasis on sustainability. This is clear in the Business Case, and also in the invitation to tender documentation (Vol. 1), which states that the Suppliers must '*work with local and national level Governments, and sector actors, to move towards developing a service delivery approach that is focused on ensuring institutional arrangements and support mechanisms for operation and maintenance over the long term, and with consideration to the full life-cycle costs of service provision (including who will bear these).*'

Sustainability is explicitly built into the programme design in so much as the PBR payment linked to outcomes incentivised the sustaining of an agreed level of results at least until programme end. However, linking prerequisites for sustainably to payments under the *Output* Phase was not explicitly required by the design (although Suppliers were free to do so). Each Supplier could choose how to promote and enable sustainability; how they did so is explored further in Sections 6.1.5 and 6.1.6 below – these sections discuss the extent to which DFID's declared commitment to sustainability was potentially marginalised by the tight timeline of the MDGs deadline and the output / outcome phasing.

4.8.3   DFID's emphasis on equity

DFID's programme design has a clear poverty focus. The overall focus of the programme was to be on '*low income and DFID priority lower middle income countries with significant water and sanitation needs*' (terms of reference, clause 1.5.). The Business Case specifies that Suppliers must '*accurately target their interventions at the poorest communities and, more specifically the poorest groups within these communities*' (Business Case, para. 121).

The programme terms of reference also require that Suppliers ensure '*that vulnerable groups are benefiting from the interventions*', and that they seek to understand '*whether there is any exclusion*' (terms of reference, clause 5.1).

The terms of reference also require Suppliers '*to have robust results monitoring and reporting methodologies and mechanisms in place … the requirements shall include [inter alia] data on number of people with access to WASH services as a result of the project, disaggregated by gender, wealth quintile, and any disability, vulnerable or disadvantaged groups.*'

4.8.4   Emphasis on equity and sustainability in Supplier activities

It was noted earlier that the Suppliers did not design their own programme-specific and country-specific theories of change to complement the generic one in the original Business Case. The processes by which activities would lead to equitable and sustainable WASH outcomes, and ultimately to health impacts, were not therefore laid out explicitly, although some of this detail is contained in the narrative proposals.

On sustainability: DFID's emphasis on sustainability was clear to Suppliers. However, as linking prerequisites for sustainably to payments under the *Output* Phase was not explicitly required by the programme design, there were variations in sustainability approaches by Suppliers. SSH4A is the only Supplier which explicitly linked certain sustainability prerequisites to its payment milestones. While most Supplier designs involved strengthening coordination at district level, for example by drawing up district WASH plans, strengthening of country coordination mechanisms at national level has not been featured during the Output Phase. It is far from certain, therefore, whether local government agencies with a permanent presence in programme districts will take on active responsibility to support and enable service functionality and use, and the practice of hygienic behaviour, beyond 2018.

On equity: While all Supplier programmes prioritised poor and under-served countries, only some implementing partners deliberately selected hard-to-reach locations, for example in Pakistan (Plan) and Kenya (Oxfam, Practical Action, Concern Worldwide). Only SSH4A (which used survey data for monitoring purposes) routinely disaggregated beneficiary results by gender and wealth quintiles. SWIFT and SAWRP, meanwhile, disaggregated beneficiaries only by gender, and adopted no clear definition of 'poor people', though there was a clear pro-poor focus in their

programme designs. As a result, DFID cannot verify to what extent people in the lowest wealth quintile have benefitted from the WASH Results Programme in the targeted locations.

It is important to note here that, while the overall programme design prioritised both equity and sustainability, these elements were not built into the *output*-level payment criteria agreed between DFID and each Supplier (note that sustainability was, however, built into *outcome* payments). In the drive to meet Output Phase targets it was therefore easy for these aspects of the programme to become marginalised in Supplier operations – even though Suppliers, like DFID, regard them as highly important. DFID was aware of these risks and one response was to introduce the tracking of measures in support of sustainability, with an RAG sustainability rating given by the MV team every quarter. The importance given to equity in hindsight is discussed in Section 7.4.1.

4.8.5   DFID's emphasis on health and non-health impacts

The logframe includes the ultimate intended impact that '*poor people benefiting from the WASH Results Programme have improved health status, as measured by reduced diarrhoea and child mortality*'. The Business Case additionally mentions a number of non-health impacts which may result from the programme, but which are not explicit in the logframe or theory of change.

It was agreed with DFID that this evaluation would not seek to measure impact, partly because the review would be completed before impacts were likely to become apparent but also because the health impacts of WASH interventions are notoriously difficult to measure. Instead, the evaluation would consider the *prospects* for impact.

Here we note some features of the programme design which could affect the prospects for health impact:

- Beneficiary numbers may be spread widely geographically, and across sub-components (water, sanitation and hygiene). DFID's design expressed no preference for geographically focused programming in order to achieve a 'critical mass' of promotional intervention or ensure that barriers to hygiene arising from severe water scarcity would be addressed.

- There was no requirement for Suppliers to work towards ODF communities, or even some minimum level of sanitation coverage at community level. While the threshold for sanitation-related health impact it is not known, it is evident that households which continue to defecate in the open pose a health risk to their neighbours, creating a strong argument for pursuing the eradication of open defecation community-wide.

- Hand-washing with soap or ash at critical times can contribute significantly to reducing diarrhoeal disease, but sector experience shows that mass-media interventions tend to have only a modest effect on behaviour, in contrast to door-to-door interventions.

- The programme allowed proposals under different lots. Lot A programmes included water[10], sanitation and hygiene, while Lot B covered sanitation and hygiene only. A limitation of the SSH4A programme under Lot B was that it offered no scope for resolving occasional situations where acute water shortages limited a community's ability to adopt hygienic practices.

These features raise concern as to whether the WASH Results Programme will have a significant impact on the incidence of diarrhoeal disease, except in locations where the programme achieved very high (>70%) community-wide latrine or toilet usage and water supply was not a constraining factor for the practice of safe sanitation and good hygiene[11]. Having said this, we acknowledge that

---

[10] Only SWIFT had a substantial water supply component which included some standalone urban water projects, while SAWRP's water component was fairly small.

[11] Note that recent sanitation RCTs in which coverage and access were only modestly increased (<70%) showed no improvement in diarrhoeal disease (Hunter and Prüss-Ustün 2016; Carter 2016).

understanding of the threshold for sanitation-related benefits, and of the possible additive effects of water, sanitation and hygiene interventions is still evolving.

4.9   MVE

At the time of programme design the MVE functions had not received the same level of attention as the Supplier-delivered programme. The procurement process for the MVE provider ran slightly later than that for the Suppliers, with the invitation to tender being issued on 22 November 2013, the tender deadline being 20 December 2013, and the selection process extending into the first few months of 2014. The contract with e-Pact was signed in May 2014. According to the MVE terms of reference, the MVE provider would 'provide an independent and rigorous MVE function for the WASH Results Programme' (MVE terms of reference).

In relation to the MV function specifically, it was stated that the provider was required to '*design and implement a system of third-party results verification (including working with Suppliers to define measurement methods and monitoring, verification and reporting requirements for payment) for those parts of projects operating under a Payment by Results framework. This will be used to verify results and outputs to inform payments. The MVE provider will audit contractors' results reporting within the contract terms related to payment by results.*' While not explicitly stated in the terms of reference, the budget available for MV services suggested that DFID was looking for a systems-based approach whereby verification was based on checking the validity of monitoring data collected by Suppliers, rather than gathering new data in the field, as sampling at a statistically significant level would not have been possible. The additional monitoring role of the MV team is described further in Section 4.10.2.

Regarding evaluation, DFID required the MVE provider to '*develop one or more impact evaluations on key issues agreed during the inception with DFID, but potentially covering one or more of the following: approaches to service delivery at scale, efficiency and effectiveness across different contexts, sustainability and Payment by Results.*' It was also expected that the evaluation would also provide learning and permit course correction.

While the Suppliers and the MVE provider each have contracts with DFID, the relationship between the Suppliers and the MVE provider is not based on a formal contract (Figure 8). A governance framework setting out key principles of engagement among all parties aims to bridge this gap. The details of the verification process are set out in Section 5.1.

Figure 8:   Triangular relationship between DFID, the MV provider and Suppliers[12]

In WASH Results, the e-Pact consortium provides the MV team, the Evaluation team, and a Learning & Dissemination function that works across both teams.
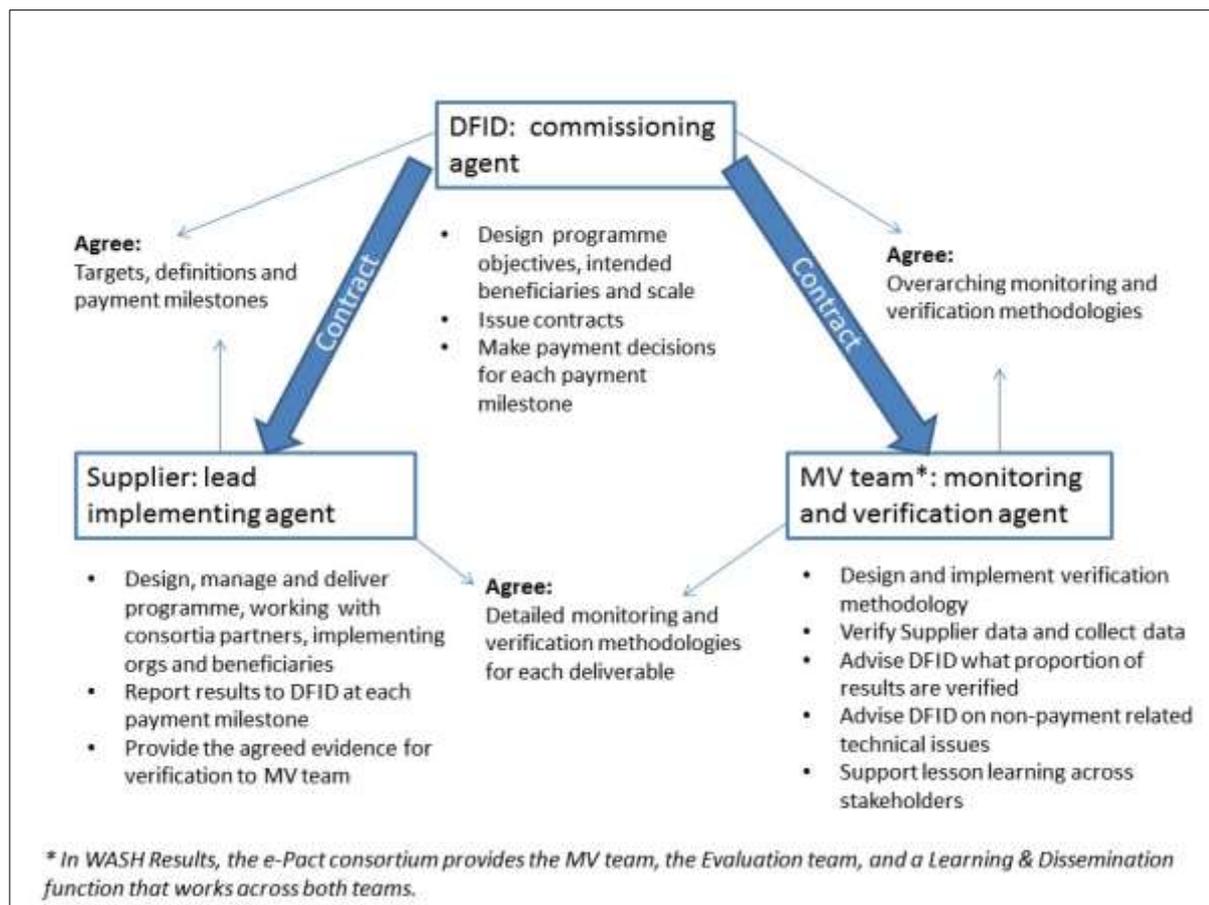
4.10   Supplier monitoring systems

4.10.1   Internal monitoring activities

DFID's terms of reference stated that Suppliers should have robust results monitoring and reporting systems which would provide the basis for verification of results, but Suppliers would be free to propose *how* they would measure outputs[13]. Knowing what would be acceptable was a challenge for Suppliers because the MV team was only appointed at programme start, hence their requirements only became apparent then and not at the proposal stage. While the MV workplan included a two-month inception period, implementation activities began immediately for Suppliers. Consequently, SWIFT and SAWRP's monitoring systems were only partially in place at the proposal stage, the intention being to fine-tune them in collaboration with the MV team. In the case of SNV, the existing SSH4A monitoring system, as used in other countries, proved to be largely fit-for-purpose, though some changes were deemed necessary by the LV.

The first few months of the programme involved intensive work by the MV team and Suppliers to establish systems that were acceptable to all parties (including the design of baseline surveys), and some further changes were made as the Output Phase unfolded – especially in Kenya, which had a wide range of partners and project types. For SAWRP and SWIFT, results data were generated based on routine progress reports from implementing partners. For many partners these databases were paper-based, though DRC used an electronic system and the WaterAid system in

---

[12] E-Pact (2016): Payment by Results – The verification approach under the WASH Results Programme (July 2016).
[13] Best practice recommends that guidance on how results will be measured and verified be available before programme implementation begins (Fritsche *et al.* 2014).

Bangladesh was mobile-based; the latter was later adopted by Plan International in the same country, though trials in Pakistan were unsuccessful. SSH4A, on the other hand, generated its output data via household surveys and government data. DFID had specified in its terms of reference that '*bidders will be required to utilise existing in-country WASH monitoring systems and to promote the development of local monitoring capacity wherever possible*'.

Designing monitoring and reporting systems and associated verification processes – particularly the indicators used as payment triggers, some of which were unique to particular sub-projects and payment milestones – took a considerable amount of time and effort for both Suppliers and the MV team, and complicated the start of programme implementation. Inevitably, the looming 2015 deadline necessitated compromises, and some baseline surveys were undertaken before MV requirements were fully developed, which posed challenges for the verification of outcome levels based on changes above this baseline.

Section 7.2.1 provides further detail on the monitoring and reporting processes planned for the Outcome Phase.

4.10.2 Monitoring activities by the MV Team

The MV team also had a monitoring role: they were intended to 'develop a monitoring framework to assess progress and performance of individual projects', to 'carry out un-announced visits to projects' and to '[monitor] risks to programme delivery and [advise] on mitigation measures', as set out in the terms of reference.

In practice, the MV team initially only tracked the progress and performance of individual projects through the verification of monitoring data, and noted any risks to programme delivery in its quarterly reports to DFID. In early 2015, DFID requested the MV team additionally monitor whether the prerequisites for sustainability were being put in place in the three Supplier programmes. The risks to sustainability were reported on in the quarterly verification reports, using an RAG rating for functional sustainability, institutional sustainability, financial sustainability, environmental sustainability and equity of use.

In this same period, DFID also requested the MV team to carry out more regular country visits to get a more detailed oversight of progress and performance of individual projects. This had budgetary implications for the MV team. Most countries were visited at latest once[14] during the Output Phase, with the MV team exploring five themes: (i) whether Suppliers were addressing sustainability aspects period; (ii) whether community engagement was adequate (to encourage sustainability); (iii) whether implementation processes were of sufficient quality to engender sustainability; (iv) whether Suppliers had allowed sufficient finance and capacity for sustainability activities post-2015; and (v) whether Suppliers were addressing environmental sustainability issues (pollution risks) through the programme. Country visits were written up as back-to-office reports. DFID also undertook three country visits during the Output Phase.

DFID stressed that this monitoring role was de-linked from the verification role: namely, that any concerns regarding sustainability would not be linked to any payment reductions.

---

[14] Three of the SNV countries could not be visited, for budgetary reasons.

5   Design of the verification system

This chapter describes the design of the verification procedures for each of the Suppliers and relates particularly to the 'relevance' aspects of the evaluation.

> **Key findings of this chapter**
>
> - The programme adopted a systems-based approach to verification. This meant that the Verifiers did not collect parallel monitoring information to assess progress, but relied on verifying the information generated by the Suppliers own reporting and monitoring systems.
>
> - The verification process has varied significantly by Supplier, being tailored to individual results frameworks. As the verification methodology was designed and adjusted while Supplier implementation activities were ongoing, several modifications were made to the evidence requirements.

5.1   Verification systems tailored to each Supplier

5.1.1   Prior experience with the PBR modality

A report by the National Audit Commission in 2015 cautioned that implementing PBR programmes requires different skills and capacities as compared to more traditional approaches used by donors and funding recipients, including strong results forecasting capability and risk analysis. It concluded that, as the market was still underdeveloped, few sector stakeholders had the required skills, driving up prices of technical assistance, lengthening start-up costs and time, and risking poor design and implementation (National Audit Office 2015). DFID also highlighted the importance of considering whether a PBR programme needs to be accompanied by technical assistance (technical support or training) or capacity building in order to facilitate the delivery of the results in question (DFID 2014a).

The team within DFID that drew up the terms of reference and managed the WASH Results Programme did not have prior experience of managing a PBR programme, and based the design on internal DFID strategy papers on PBR and inputs from the PBR team in DFID. The Suppliers also had limited prior experience of designing and implementing PBR programmes:

- SAWRP chose Plan International to lead their consortium as Plan International had prior experience of implementing PBR programmes in education (the Girls Education Challenge Fund), whereas for WaterAid it was their first PBR programme and they considered it a learning opportunity.

- Oxfam chose its SWIFT consortium partners before it became clear in 2013 that the programme would have a PBR funding modality. Partnerships were not changed once the PBR modality was announced, even though none of the partners had PBR experience.

- SNV had not implemented a PBR programme in WASH before but they adapted their existing SSH4A design to suit a PBR payment modality.

The verification team members also did not have prior experience of implementing a verification process, though some team members had evaluated other PBR programmes within and outside the WASH sector, but they drew on expertise within e-Pact, which had designed a verification process for government-run health facilities in Pakistan. The LVs were all experienced WASH experts who were well placed to assess the quality and sustainability of WASH interventions.

5.1.2   Verification framework

Due to the PBR financing modality of the programme, Suppliers are 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** in this programme. 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-generated data:** in parallel to the desk-based verification of evidence described above, the MV team carry out field visits to double-check the veracity of evidence submitted and the quality of results achieved, and to assess the likely sustainability of result 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.

Table 8: **Durations of verification cycle for each Supplier**

SWIFT	SAWRP	SSH4A
• Several weeks for SWIFT to provide evidence • From evidence submission: two weeks for clarifications on evidence • 2.5 weeks (in parallel with above) for country report submission • 1.5 weeks for Supplier report submission and quality assurance • Verification report was submitted one week before payment meeting	• One week for SAWRP to provide the draft database for sampling • Five working days for provision of sample secondary evidence to Country Verifiers • 2.5 weeks (approx.) to verify evidence, ask for clarifications and write verification reports • Verification report was submitted 10 days before the payment meeting	• Two weeks for field spot checks *(when required)* • Two weeks for the county verification reports • One week for the LV to draft the verification report • Verification report was submitted one week before payment meeting
Five- to six-week cycle	**Five-week cycle**	**Three- or five-week cycle**

This process was repeated almost every quarter under the Output Phase, and has been repeated every second or fourth quarter during the Outcome Phase to date, depending on the Supplier (see table below). The precise dates were agreed in advance each quarter, to allow for festivals, holidays, et cetera.

Table 9: **Timing of verification rounds for each Supplier**

	Output Phase							Output Phase							
	2014			2015				2016				2017				2018
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
SWIFT	●	●	●	●	●	●	●	●				●				●
SAWRP		●	●	●	●	●	●		●		●		●		●	
SSH4A		●●	●		●	●	●				●		●		●	●

In addition to this verification function, the verification team also have **an advisory function**: they provide advice to DFID on the quality of Supplier interventions – in particular, the likely sustainability of the WASH outcomes achieved through the programme. The advisory function aims to flag up quality issues, to assess where there is room for improvement for the Supplier in regard to delivering WASH services. This advice does not directly affect payment but if key risks are identified, action plans are agreed with Suppliers through After-Action Reviews that set out mitigation measures.

The verification team comprises one LV, one Deputy Verifier (previously called 'Shadow Verifier' (SV)) and a series of national Country Verifiers for each Supplier (see Figure 9). The relative level of effort required of different members of the verification team is tailored to the tasks involved in each verification round (see Table 10).

Figure 9:   Organogram of the MV team
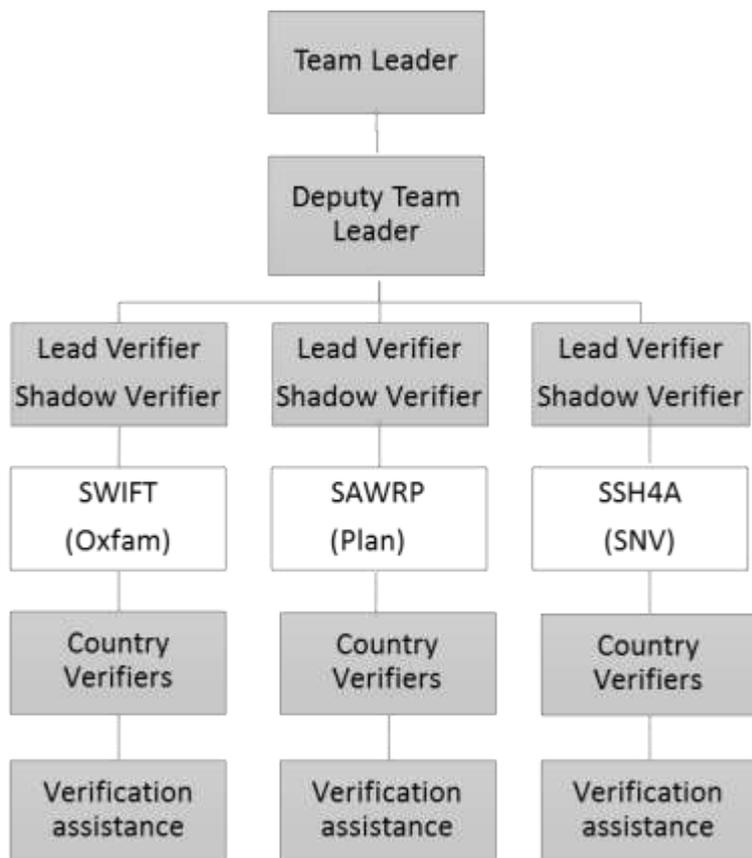
Table 10:   Variations in the effort required for different verification rounds

Supplier	Days for process deliverables	Days for output deliverables
SWIFT	Four days for LV, six days for SV Eight days for each Country Verifier	Eight days for LV, five days for SV (+ six days for each country visit) 10–12 days for each Country Verifier
SAWRP	n/a	Eight days for LV, five days for SV (+ eight days for each country visit) 8–15 days for each Country Verifier
SSH4A	Eight days for LV, three days for SV Five to nine days for each Country Verifier *(depending on whether field spot checks were done, which were only made by four to five Country Verifiers when required)*	10 days for LV, four days for SV (+ eight days for each country visit) Five to nine days for each Country Verifier *(depending on whether field spot checks were done)*

5.1.3   Verification process linked to internal quality assurance procedures

The verification process was largely built around Suppliers' existing M&E frameworks. However, it also built on their internal quality assurance (QA) procedures where possible, and appraising these allowed verifiers to check the reliability of reported data. Internal QA procedures varied by Supplier and included, for example, mechanisms to ensure that minimum standards for water-points and latrines were met, and internal checks, from local to global level, on progress reports submitted from a lower organisational tier.

5.1.4   Tailoring evidence requirements to each result

The PBR literature indicates that third-party independent verification should be sufficiently 'light' to minimise verification costs and the reporting burden for Suppliers, but sufficiently 'heavy' to avoid gaming and manipulation (Clist and Verschoor, 2014). The evidence requirements for verification to take place should be clear, feasible (the verification cycle timeframe is consistent with contractual obligations and suppliers' cash flow constraints) and appropriate to the type of result being verified, to reflect the realities of the specific organisations and processes involved (Chowdry, 2011; Witter *et al.*, 2012).

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 which pertain 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 implementing partners. This resulted in 42 different Form 2s for the Output Phase (see Table 11), with a separate set of Form 2s currently being designed for the Outcome 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.

Table 11:   Variations in the number of Output Phase Form 2s by Supplier

Supplier	Number of Form 2s	Description
SWIFT	30	Forms differed for water, sanitation, and hygiene, all of which had separate forms for intermediate results, early sustainability systems and outputs (3x3 forms). These forms varied widely across the five different types of implementing partners in DRC and Kenya (two partner groupings did water/sanitation/hygiene, one did sanitation/hygiene, one did only hygiene and one only water)
SAWRP	8	One for water, one for sanitation, one for hygiene in communities and one for hygiene work in schools (Unilever only) – for Bangladesh and Pakistan respectively (4x2)
SSH4A	13	13 Form 2s across six results packages (including one for sanitation and one for hygiene)
Total	42	

As each Supplier was assigned a different LV, these forms evolved relatively independently to suit the evidence requirements of different deliverables, and to match the different reporting and monitoring data available for different Suppliers.

The minimum level of evidence required was agreed with the Supplier for each deliverable through a process of negotiation, where the Verifier suggested items of evidence and the Supplier checked the feasibility of providing this information within the required timeframe. This approach was

important for maintaining good relationships between Verifiers and Suppliers, and allowed enough flexibility for adjustments to circumstances to be made.

- **SAWRP** was the only consortium where every verification round focused on the same deliverable: numbers of people reached with WASH services. This straightforward design lent itself to drawing up one single evidence database for the Output Phase, covering all water-points and latrines constructed by partners, plus school children and other people reached with hygiene promotion. This database was submitted to the verification team every quarter, and from it a number of water-points, latrines, and hand-washing beneficiary villages and schools were sampled using a Lot Quality Assurance Sampling (LQAS)[15] method to check the supporting evidence. It took several verification rounds to agree on the precise specification of the supporting evidence required and there were challenges in ensuring that the same facilities were not counted as new in successive reports.

- For **SWIFT,** each verification round focused on three types of 'result': intermediate results, early sustainability systems and output delivery. What was actually measured varied from one project to another and over time, and many of the 'results' were in fact implementation processes. For these, the Verifier asked Suppliers to provide a sample of evidence, as opposed to drawing a sample him/herself, as it was not feasible to sample across such a large number of partners. One area of disagreement between Supplier and Verifier related to the multiplier used for counting sanitation beneficiaries in one project location, as it became clear over time that the average number of users per household toilet was going down.

- The **SSH4A** contract contained a clearly defined set of results divided into 'results packages'. These packages changed from one quarter to the next over most of the Output Phase. Evidence for process milestones was mostly derived from household surveys, the content of which was based on SSH4A's previous work in Asia.

The process of meeting evidence requirements was relatively straightforward for SSH4A because it adopted an established monitoring framework familiar to SNV and applied it across all nine countries. Furthermore, SSH4A has been implemented by a single organisation, not a consortium, enabling SNV to standardise reporting requirements across the country teams. In contrast, both SWIFT and SAWRP are consortia with multiple partners and are led by managers who are not fully familiar with the detail of each organisation's systems. In Kenya, much time was spent identifying the best compromise between rigorous but pragmatic verification systems, though in DRC bespoke systems were set up at the outset – a significant investment by the Supplier, which proved valuable in the longer term.

5.1.5   Tailoring systems appraisals to Suppliers

Systems appraisals are intended to check whether internal monitoring and reporting systems are robust enough to generate the evidence needed for verification. The same appraisal format was used for all Suppliers, but the checks were carried out in a different fashion by each Verifier:

- For SAWRP, the Verifier first undertook a desk review of the systems in place, and then repeated this check for several quarters, in the form of a 'systems monitoring' to check if the intended procedures were being used as planned by all implementing partners.

- As SWIFT consisted of many different partners, each with their own internal monitoring and reporting systems, the Verifier first checked the processes in place at HQ, did the same at country management level the following quarter, and in the third quarter did so at partner level.

---

[15] LQAS, sometimes also called 'acceptance sampling', involves taking a small random sample from each set of items in the population, and testing each sampled item to determine whether it meets a predetermined standard of quality. LQAS provides less information than representative sampling but often requires substantially smaller sample sizes.

While this set-up allowed the systems appraisal to be more manageable every round, it also meant that if any issues were discovered at partner level then three rounds of implementation and verification had already passed.

- As the results packages for SSH4A differed for every milestone, a distinct systems appraisal was undertaken at the start of each verification round to check whether the procedures for those specific evidence requirements were in place.

The main purpose of, and challenge in, undertaking systems appraisals was what action to take if the system was not deemed robust enough to generate the required evidence. The MV team kept a log of the outstanding issues and discussed these in the After-Action Review. In the case of SAWRP, outstanding issues were listed in the verification report and were helpfully classified with an RAG rating to facilitate decision-making. It was DFID's role to take a decision on which recommendations the Supplier should take forward, and when. This allowed a clear separation between the advisory role of the MV team and the executor role of DFID.

However, agreeing on which recommendations to implement, and on the timeframe, was sometimes challenging, as these often had cost implications for Suppliers. The majority of recommendations were taken on board, which led to improvements being made to Suppliers' monitoring systems (see Section 6.2.2). However, some recommendations of the MV team were not implemented by Suppliers, predominantly due to cost reasons. The conducting of the baseline surveys is a key example of this: the MV team recommended larger, more statistically robust, baseline surveys, but this recommendation was only partially followed by Suppliers. This will pose risks for the reliability of the outcome surveys based on these baseline findings.

5.1.6   Country Verifier field visits

Originally, the systems-based approach to verification was only designed around desk-based review of documentation, as set out in the tender document. However, during the inception phase, the MV team modified this to include field visits by a Country Verifier, who would provide an extra layer of QA in order to check whether what was reported had actually taken place. The visits formed part of the 'monitoring' role which had been assigned to the MV team (see Section 4.10.2).

Due to budget constraints, these field visits by the Country Verifier could not cover a statistically representative sample of the results reported. Instead, the purpose was to gain a qualitative impression of whether reported results took place as intended, and also to explore issues relating to quality.

Initially, field visits by the Country Verifier were scheduled to take place after evidence had been submitted by Suppliers, in order to check a sample of that evidence. In practice, these visits were modified in a slightly different way for each Supplier, to allow more time for evidence submission under each verification round:

- **For SAWRP**, field visits took place prior to evidence submission. Visits were not used to check the quality of water-points, latrines, or BCC sessions, or whether reported activities had actually taken place, but to confirm whether monitoring and reporting processes were being implemented as intended. During the field visits the field activities were also reviewed and the Country Verifiers provided feedback on their quality. The field visits were largely decoupled from the broader verification cycle.
- **For SWIFT**, the field visits were decoupled entirely from the verification cycle and were scheduled to take place after the verification report had been submitted. This approach reduced the pressure on partners and allowed any insights from field visits to feed into the next

verification cycle instead. Field visits focused particularly on flagging any quality and sustainability issues arising as part of implementation.

- **For SSH4A**, around half of the nine countries received a field visit by the Country Verifier for each verification round; the possibility of additional scrutiny was deemed sufficient to incentivise those that missed out on a country visit in any given round. A desk-based check of all documentation was undertaken for every country in every round, and if this check did not reveal any issues then a field visit was not necessarily undertaken. Field visits, where they were made, could have a direct bearing on verification decisions, and hence Supplier payment.

5.1.7   Consequences of 'failing' a verification round

While the verification process was designed as a step-by-step process, all three LVs explained that in practice the process was iterative, giving Suppliers the opportunity to comment on evidence requirements and submit additional evidence later within the same verification quarter. This approach was found to be necessary in order to make a complex system workable and fair. LVs felt that without this flexibility, far fewer results packages would have passed each verification cycle.

Any payment reductions resulting from 'failing' a verification requirement were deliberately very small, acting as warning sign:

- For SAWRP, if evidence was found to be insufficient then a proportional reduction in payment was made for that sub-district. Any deduction was intended as an incentive for partners to submit better evidence in the next round. In practice, evidence was found to be adequate in most cases. Occasional reductions were very small, and full payment was still received by the end of the Output Phase as all partners exceeded their targets and therefore caught up on payments. This risk management strategy had been deliberately built into the SAWRP proposal.
- For the process-based results, SSH4A opted for an arrangement whereby verification results were not proportional but expressed as a simple pass or fail. This approach was initially also applied to the output milestones. However, as a result of over half the evidence being deemed insufficient during SNV's first hygiene output result, the evidence requirements were adjusted and the verification was switched to a proportional process, allowing SNV to re-submit the missing evidence in later quarters. In later surveys, any reductions in payment were only applied to the survey sample under examination, with only very small payment reductions possible.

For both SAWRP and SWIFT, partners opted to retract some payment claims during verification quarters because their internal QA check had indicated that they were missing certain bits of required evidence. While this may have caused some cash flow issues, it enabled full payment to be received in the next quarter. SSH4A also chose to retract the sanitation survey results for South Sudan and Ghana for the same reason – because internal QA checks had flagged missing bits of information.

Overall, the quality of the evidence submitted improved over time, as evidenced by the fact that fewer payment reductions needed to be made in later verification quarters. In addition, Suppliers also reported that their monitoring systems were strengthened as a result of the scrutiny of the verification process (see Section 6.2.2).

6   Implementation during the Output Phase

Having described and analysed the programme *designs* of DFID and the Suppliers in Chapter 4, this chapter presents and analyses the experience of implementation under the Output Phase. It focuses on the question of which factors influenced the effectiveness, efficiency and likely sustainability of the activities implemented to date.

<div style="background-color:#6b9ca0; color:white; padding:4px;">

Key findings of this chapter

</div>

- The vast majority of output targets were met, with significant over-achievement in many countries.

- Suppliers set conservative output targets, given the short time available for the Output Phase and given the fact that the PBR contract by design did not reward over-achievement but penalised under-achievement.

- Suppliers expanded, or built upon, earlier WASH programmes, implementing mostly tried and tested approaches, except in Kenya where some SWIFT partners implemented approaches that were new to them, or that were implemented in a new area. PBR risks and tight timelines discouraged innovation and encouraged Suppliers to adopt (mostly) familiar, tested approaches. In both cases, programme management was flexible and responsive to changes on the ground, with many examples of results-oriented problem-solving. On the one hand, by its non-prescriptive nature, PBR allowed country teams the flexibility to adjust implementation approaches where needed. On the other hand, the time pressure of the MDGs deadline allowed little time for reflection and adjustment.

- The key factors affecting progress on implementation included: (a) the PBR modality of payment (including the requirement for rigorous independent verification); (b) the strict deadline for delivery of beneficiary number results by December 2015; and (c) the programme 'shape' or phasing. These factors focused minds on results, increased the scrutiny of the monitoring data and affected the risk management strategies of Suppliers.

- While all Suppliers were committed to delivering sustainable WASH services, the abbreviated nature of the Output Phase, and the focus on quantifiable results, made it difficult for some SWIFT and SAWRP partners to give adequate attention to achievement of sustainable outcomes. In other cases, however, SWIFT and SAWRP partners prioritised sustainability despite these perceived disincentives. SSH4A is the only Supplier that explicitly built the prerequisites for sustainability into its payment milestones and that laid the groundwork early on to enable sustainability, with most sanitation marketing, governance and BCC activities starting from the outset.

- The pressure of both the PBR modality and the MDGs deadline to deliver results discouraged implementation through government for some Suppliers, especially in fragile-state contexts where government capacity was too weak to deliver activities under a tight timeline.

- While most Suppliers did not pass PBR risks down to local partners, the PBR modality still changed the nature of the relationships with local NGO and government partners to some extent, in some cases placing considerable strain on relationships.

- The verification process underlying the PBR modality clearly led to higher scrutiny of the Suppliers' monitoring data. This required a higher monitoring effort by Suppliers; more than had been expected at tender stage, with a significant reporting burden experienced by staff, particularly for process-based milestones. However, it also led to the strengthening of monitoring systems, improving the frequency and reliability of monitoring data and allowing more regular interactions between M&E and technical teams.

- The PBR modality and the MDGs deadline affected the risk management approach of Suppliers: how they managed their consortia, the extent to which they passed risks down to local partners and how they scheduled payment timings.

- It is unclear whether PBR has enhanced VFM, but it is evident that not all Suppliers were able to price in risks adequately in their budget proposals.

- While Supplier programmes included some internal learning initiatives, the tight timeline left little room for this.

Out of the 10 programme features highlighted in Section 4.2, three are arguably the strongest determinants of how the programme has played out so far. These are: (a) the PBR modality of payment (including the requirement for rigorous independent verification); (b) the extremely tight and strict deadline for delivery of beneficiary number results by December 2015; and (c) the programme 'shape' or phasing, involving Output and Outcome periods. Their individual effects are tightly intertwined. Figure 10 below provides a diagrammatic representation of the resulting impact pathways, which are described in this chapter. The three aforementioned factors resulted in increased scrutiny of the monitoring data, focused minds on results, and affected the risk management strategies of Suppliers. Box 3 below provides further detail on the theory behind PBR incentives.

Figure 10:   Examples of impact pathways
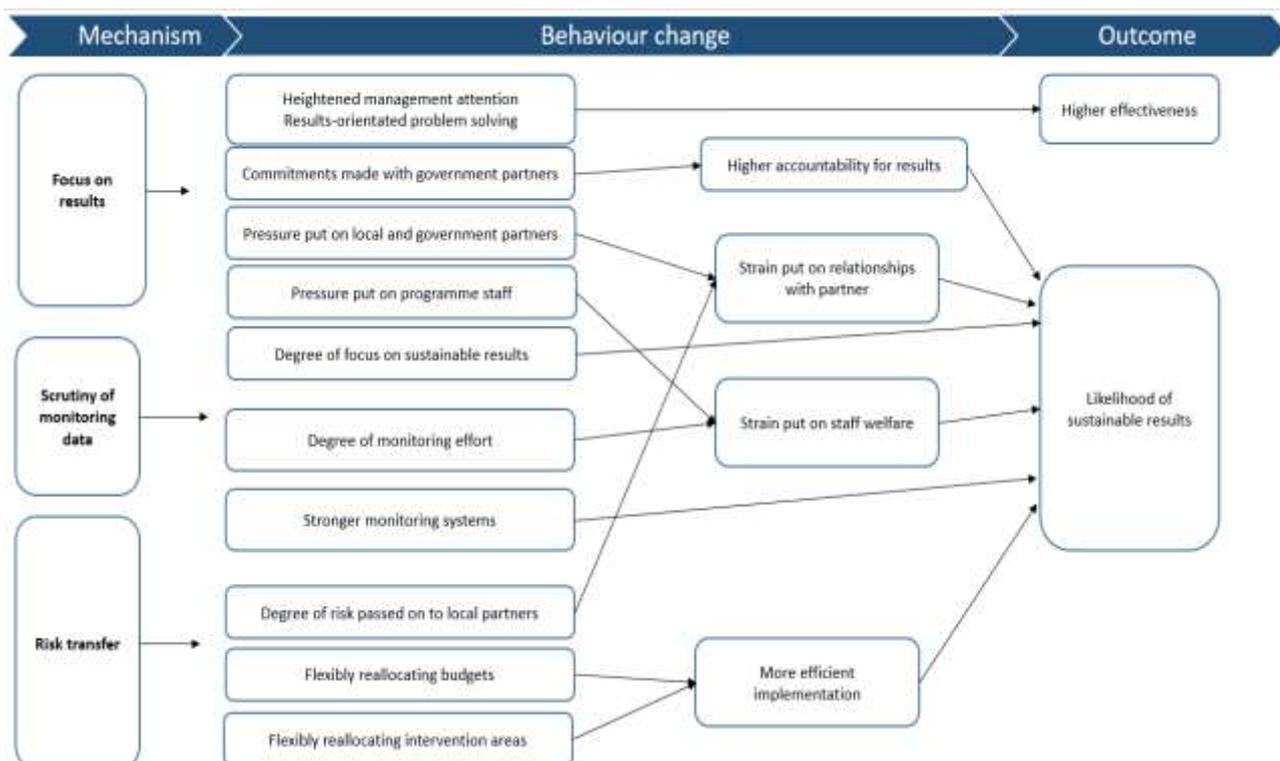
Note that different changes in behaviour could have positive or negative effects on outcomes

Box 3: Theory behind PBR incentives

The underlying rationale for PBR assumes that Suppliers have systems that are capable of responding to, and acting on, the incentives created under a PBR modality (Perrin 2013). It is also assumed that implementation processes are sufficiently flexible that they can be modified to optimise delivery, within the timeframe of each verification cycle. By not specifying *how* results should be achieved, it is assumed that Suppliers can adjust implementation processes to improve outcomes (Birdsall and Savedoff 2010, Bertone and Meesen 2013). Furthermore, there is an implicit assumption that Suppliers have sufficient skills and capacity to monitor progress sufficiently accurately and sufficiently frequently to identify trends in performance and identify the underlying drivers, and that they have effective management systems in place which can adjust implementation processes in response to this learning (improved performance management; Birdsall and Savedoff 2010, Bertone and Meesen 2013).

6.1   'Focusing minds on results'

It is clear that both the PBR modality and the MDGs deadline of December 2015 'focused minds on results'. This had several implications for the implementation approach chosen, the numbers of results achieved, and the quality of these results. These are discussed in turn below.

6.1.1   Vast majority of output targets achieved

The three Supplier programmes were quite different in terms of their organisational make-up, the range of activities undertaken, the number of countries involved, and the extent to which they were applying tested operational approaches or using implementation models that needed some fine-tuning. For these reasons, it is difficult to say that one of the programmes performed significantly better or worse than another, but they all met their contractual targets. While DFID offered a three-month extension to Suppliers from January to March 2016 to lessen the pressure of the MDGs deadline, only SWIFT took advantage of this extension because it had faced more delays as a result of its implementation through government (see Section 6.1.6).

The Suppliers have in fact reported over-achievement in most cases (Table 12). In the case of SNV, for example, five out of the nine participating country offices have reported that they exceeded their targets by more than 2.5 times the original target. One reason for this over-achievement was that Suppliers had set conservative targets, as a result of the PBR modality, the tight MDGs deadline, and to match to the capacity of local partners (see Section 4.3.3). Secondly, both SAWRP and SSH4A explained that the internal targets agreed with their partners in country were 10%–30% higher than the targets agreed with DFID. This was a risk management strategy chosen to provide a buffer, so that this over-achievement could compensate any under-achievement which might arise in some cases where unexpected challenges were faced (see Section 6.3.1). Having such flexibility was cited as a benefit of PBR, though there were also limits to this flexibility (see Section 6.3.4).

DFID indicated that grant programmes implemented by NGOs sometimes propose targets which are later not met, explaining that having greater certainly in regard to results delivery was one of the main expected benefits of PBR (DFID 2015a). Under this programme, it appears that the PBR modality encouraged careful reflection of which level of delivery was feasible under a tight timeframe. This was reinforced by the fact that PBR payments were linked not just to outputs but also to outcomes having to be sustained until programme end. Several NGO partners explained that, at the design stage, they reduced their targets once they learned about the sustainability requirements that were linked to the PBR modality.

> *I don't think PBR makes us less ambitious but that it made us more realistic in the probable numbers of people to be reached. More careful analysis of the baseline situation was done. I think that is a good thing because under grants, we often commit to ambitious numbers which are hardly met.* – SSH4A staff member

> *We based our decision of scale on how quickly our partners could scale up and on whether we could sustain that level of capacity after the programme had closed (for the next 10 years). If we were only interested in delivering the [beneficiary] numbers until December 2015, we could have doubled the number of beneficiaries that we could reach.* – SAWRP staff member

> *PBR makes you budget more carefully; which activities can I actually deliver with this money?* – SWIFT NGO partner

Table 12:   Output targets and results per Supplier programme

	Water supply access			Sanitation access			Reached by hygiene promotion			Hygiene in schools		
	Target	Verified result	%	Target	Verified result	%	Target	Verified result	%	Target	Verified result	%
SWIFT												
Kenya	201,300	186,389	93%	30,900	24,733	80%	330,000	329,035	100%			
DRC	647,067	492,496	76%	428,914	423,280	99%	697,257	645,664	93%			
Total SWIFT	**848,367**	**678,885**	**80%**	**459,814**	**448,013**	**97%**	**1,027,257**	**974,699**	**95%**			
SAWRP												
Bangladesh	137,750	170,260	124%	644,176	745,591	116%	665,570	765,169	115%	4,158,000	4,322,807	104%
Pakistan	90,000	111,662	124%	1,000,000	1,202,530	120%	1,000,000	1,164,035	116%	786,000	787,408	100%
Total SAWRP	**227,750**	**281,922**	**124%**	**1,644,176**	**1,948,121**	**119%**	**1,665,570**	**1,929,204**	**116%**	**4,944,000**	**5,110,215**	**103%**
SSH4A												
Ethiopia				194,000	143,220	74%	272,000	542,115	199%			
Ghana				270,000	106,117	39%	272,000	454,309	167%			
Kenya				240,000	364,346	152%	417,000	1,082,977	260%			
Mozambique				200,000	263,064	132%	272,000	1,340,422	493%			
Nepal				240,000	294,801	123%	272,000	504,274	185%			
South Sudan				190,000	*TBC*	-	272,000	TBC	-			
Tanzania				190,000	367,183	193%	400,000	1,108,016	277%			
Uganda				330,000	504,015	153%	303,000	2,207,597	729%			
Zambia				230,000	284,051	124%	250,000	587,100	235%			
Total SSH4A (excl. South Sudan)				**1,384,000**	**2,326,797**	**168%**	**1,914,000**	**7,826,810**	**409%**			

Note: The SWIFT consortium has reportedly delivered more results than are shown in this table, but only results related to meeting the output targets were submitted for verification.

6.1.2   Most Suppliers built on earlier WASH programmes

Most Suppliers expanded, or built upon, earlier WASH programmes, and could therefore utilise tested operational approaches and established institutional relationships with local government agencies and NGOs/community-based organisations. Adopting tested models and selecting a limited range of projects and partners enabled these Suppliers to achieve economies of scale. SAWRP also operated at scale, and had the unique feature of a dedicated and largely autonomous school hygiene promotion project operated by Unilever. SNV also sought to deliver activities at scale in partnership with government partners, operating either district-wide or in a selection of sub-counties. SWIFT, in DRC, also operated at scale, covering several districts under the national VEA programme.

The only exception across the 12 countries was SWIFT's operations in Kenya, which, arguably, were not so much a coherent single programme as a package of small- to medium-sized projects that operated largely autonomously, though Oxfam, as consortium manager, did much to ensure that each partner was on track and that there was good communication between the various INGO partners. Irrespective of PBR monitoring and reporting requirements, the transaction costs of management and coordination in Kenya have been particularly high. In addition, the programme elements included operational approaches which were not all tried and tested at the outset. In the case of rural sanitation and hygiene promotion the programme struggled to gain traction, especially in Turkana.

6.1.3   Heightened management attention

Across the three case studies, Supplier behaviour was affected by an awareness that the WASH Results Programme was a new departure for the sector and would attract much external attention. The programme was therefore afforded special status, with the Suppliers very conscious of reputational risk and going to great lengths to ensure that targets were met. In Kenya, for example, there were several occasions on which implementing partners drew on alternative sources of funding to resolve obstacles to progress: Oxfam paid for a power supply to a borehole when the utility failed to do so (this was the utility's responsibility under a cost-sharing agreement); Sanergy hired a consulting firm to help them re-develop their business model; and BBC Media Action paid for a legal power connection to a community radio station, which made hygiene promotion broadcasts for the programme.

Other behaviours were, however, less beneficial: Oxfam management, for example, waived some of its procurement rules to speed up internal processes. Pressure was also put on staff (see Section 6.1.8) and on local partners (see Section 6.1.7) to speed up implementation.

Not all these examples of the programme being treated as a special case can be regarded as desirable in the longer term. Once PBR becomes more widely used, as seems likely, heightened management attention can no longer be afforded to every PBR programme.

6.1.4   Results-oriented problem-solving and learning

While the Business Case set out to encourage innovative WASH interventions, most country teams implemented mostly familiar, tried and tested approaches, to reduce the risk of not meeting targets. The country teams interviewed explained that in order to minimise the risks of non-payment under the PBR modality, there was **reticence about innovating**. The only examples of innovation mentioned by Suppliers across 12 programme countries were the use of

solar-powered water pumps in Kenya and mobile-based monitoring by SAWRP. Both, however, are not strictly 'innovative', in the sense that they are both technologies which are used by WASH practitioners elsewhere.

However, programme management was flexible and responsive to changes on the ground, with many examples of results-oriented problem-solving[16]. While not necessarily resulting in genuine innovations, this rapid testing and feedback process allowed implementation activities to be adjusted to make them more effective, thus ensuring targets were met. Suppliers interviewed during country visits felt that PBR made Suppliers ensure that they adjusted their approaches to ensure targets were met. While it is in line with best practice to observe and adjust implementation approaches in response to learning, conventional grant projects may not necessarily place a strong emphasis on iterative learning and/or may have too prescriptive an implementation approach to allow much adaptation and learning.

> *PBR puts the spotlight on causality, encouraging attention to be given to 'how' planned activities will lead to outputs and to outcomes.* – SAWRP staff member

---

Box 4: Examples of results-oriented problem-solving and learning

- SSH4A staff felt that PBR made programme staff repeatedly adjust implementation to ensure targets were met. The Uganda team deliberately adopted a phased approach, first piloting a given approach, allowing rapid feedback and then adjusting activities where needed. The Nepal team modified the CLTS approach to the socioeconomic context of different castes. The Tanzania team modified their behaviour change messaging to focus especially on the barriers to latrine usage, as the government was enforcing latrine construction but not doing enough to encourage usage.

- In Pakistan, SAWRP local partners spent more time in communities due to stringent monitoring requirements. This allowed them to respond to local issues faster. A similar experience of community engagement allowing more responsive programming was reported by BBC Media Action, a partner of SWIFT in Kenya.

- In response to the unreliable power supply and high price of fuel in remote parts of northern Kenya, SWIFT teams adjusted their usual approach to implement solar-powered water supply systems under this programme.

For other teams, the time pressure of the MDGs deadline to some extent cancelled out the benefits of having a flexible PBR programme which was non-prescriptive in its implementation approach:

- **Several NGO partners explained that taking stock of lessons could only happen *after* completion of the Output Phase.** Adjusting implementation approaches until they proved effective was not always possible within the short timeframe of the Output Phase. SNV also found the implementation in the fragile-state context of South Sudan challenging, was not able reach its targets, and decided in September 2016 to discontinue programme activities in South Sudan. This experience provides some insight into how the pressure to deliver results under a PBR framework may be particularly challenging in fragile-state contexts, where delivery is already significantly affected by factors outside of the programme's control, though further interviews are needed with the SNV teams to explore this aspect in more depth.

- **The urgency of the MDGs deadline meant that some local partners were not able to undertake a detailed baseline assessment, in order to understand the socioeconomic, political and environmental aspects which may pose challenges to WASH implementation**. For SAWRP in Pakistan, for example, baseline sanitation coverage in the target areas was found to be higher than anticipated, requiring local partners to increase their geographical scope to meet targets – which raised implementation costs due to the need for

---

[16] This term is mentioned by the Swedish International Development Cooperation Agency (Sida) (2015).

larger field teams, transport provision and other support. Conversely, in northern Kenya, local partners flagged up the fact that better understanding of the baseline context would have helped them to design different implementation approaches.

Overall, it is important to stress that, while PBR encouraged adaptive programming for some teams, this can also be achieved under other funding arrangements, provided a responsive management framework is in place.

6.1.5   Focus on sustainable WASH outcomes

There are significant differences between how SWIFT and SAWRP, on the one hand, and SSH4A, on the other, approach sustainability. These are discussed in turn below.

While all Suppliers were committed to delivering sustainable WASH services, the abbreviated nature of the Output Phase, and the focus on quantifiable results, made it difficult for some SWIFT and SAWRP partners to give adequate attention to sustainability; there were some cases where the quantity of results was given priority over quality (see box below).

Box 5: Examples of cases where a sustainability focus was 'crowded out'
• Local partners in Pakistan explained that they delayed the majority of their capacity building inputs for communities and government until the Outcome Phase. • In rural Kenya, while Water Users' Associations were formed under SWIFT and some basic training was provided for pump operators, little more was done by the end of the Output Phase to support and enable the long-term use and functionality of water supply schemes. One respondent observed that implementation had felt like an emergency project. • Some SAWRP partners reported that efforts were so focused on achieving outputs that they lost sight of the link to outcomes (and the outcome strategy was only designed in mid-2015). Some local partners in Pakistan also reduced the time spent triggering each community, which could make it more difficult to secure lasting behaviour change.

In contrast to this there were also good examples of cases where partners prioritised initiatives in support of sustainability despite the time pressure in the first half of the programme; see Box 6.

Box 6: Examples from SWIFT in Kenya of strong focus on sustainability
• Oxfam Kenya used high quality water supply pipes and tanks made from galvanised steel, and they used hybrid power schemes based on both solar and standard power sources to minimise interruptions to the supply. The Oxfam team also recognised the need to complement infrastructure support to utilities with institutional strengthening in the Outcome Phase. • Similarly, WSUP in Kenya established a productive working relationship with Nairobi City Water and Sewerage Company (NCWSC) that addressed not only the technical dimensions of improving water supplies but also the need to engage residents and landlords effectively in the initiative. • Sanergy in Kenya did not push fresh-life operators to take on more toilets than they could successfully handle. • BBCMA in Kenya did not opt to work with larger radio stations which could reach more people through hygiene promotion broadcasts, because its ethos is to work with small community-based radio stations to develop their capacity for the longer term.

While the global programme design did not directly link sustainability measures to payment criteria in the Output Phase, SSH4A made its own decision to design its payment milestones to include

sustainability components. Drawing up district sanitation plans, for example, was an explicit deliverable; sanitation marketing was also introduced at an early stage. This is encouraging, though insights from the Uganda country visit indicate that that long-term sustainability in that country is heavily dependent on continued support by local government, which currently has low capacity at district and sub-county level.

6.1.6   Focus on governance

Overall, all Supplier consortia placed a strong emphasis on working in partnership with government and aligned their activities with national WASH policies where possible, most likely due to their organisational commitment to sustainability. The implementation model for SAWRP in Pakistan, for example, was closely aligned with the national sanitation strategy, PATS. Under SWIFT, government was consulted on activity design[17] and participated to some extent in implementation and monitoring work, bearing in mind, however, the under-resourced nature of many government departments, especially in post-conflict DRC.

Under SSH4A, SNV placed a strong emphasis on alignment with government by linking their sanitation and hygiene targets to government district plans, and they involved government staff in the monitoring of progress against these targets. This strengthened the incentive for SNV to achieve these targets as they wanted to honour their commitment to the government. In Ethiopia, SNV aims to continue supporting the government even though the output targets have now been met. This example shows that in the case of SNV the choice to involve government in sanitation and hygiene commitments also had the effect of making results more visible, and thereby leading to **increased accountability** over the agreed targets. The PBR literature supports this notion, showing that public reporting of results will strengthen incentives to achieve results under a PBR contract (Lindenenauer *et al.*, 2007).

However, the extent to which the programmes were implemented directly through local government partners varied. The urgency of results delivery affected the choice of implementing partners, with several examples from SSH4A and SWIFT of cases where the implementation modus was switched from collaboration with government partners to implementation directly through NGO partners (see box below). This was done both in recognition of government capacity constraints, which could have caused delays in delivery, and a need to gain greater control over implementation activities, with scope to adjust them when needed (see Section 6.1.4).

A key counter-example is SWIFT's work with urban water utilities in Kenya. Here activities were implemented directly through the public water utility. This, however, reduced the Supplier's control over the timing of key processes, such as procurement. Local partners reported that having to closely supervise the utility put a strain on their relationship (see Section 6.1.7). One water supply and sewerage project had to be cancelled when it became clear that it could not be delivered in the time available, but by this time the implementing partner had incurred considerable expenses on design and other preparatory work.

In those countries where implementation was at least partially done through local government staff, this was achievable thanks to the following approaches:

- having an established relationship with national and local government agencies;
- formulating joint plans (in some cases a **district or county sanitation plan**), with defined tasks and responsibilities for each party; and

---

[17] SWIFT negotiated a partial deviation from national strategy in northern Kenya, as Oxfam introduced subsidised latrine slabs.

- supporting logistical expenses and allowances for government staff.

Box 7: Examples of cases where the focus on results affected the choice of implementing partners

- While SNV aimed to implement through local government, reduced government staffing levels and capacity constrains in rural districts posed challenges in Ethiopia, Kenya and Uganda. In Kenya, SNV chose to employ field monitors to supplement the efforts of government staff. In Mozambique, local NGOs (LCBs) supported government CLTS facilitators. In Uganda and Zambia, SNV chose to implement entirely through local NGOs (LCBs), though there was still minor government involvement in CLTS triggering and follow-up activities. Hygiene promotion was also largely NGO-led, though there was some involvement of health assistants. In the other SSH4A countries, SNV implementation through local government was only possible by placing considerable pressure on these partners, which affected their relationship (see Section 6.1.7).

- While WaterAid also usually implements through local government partners in Pakistan, under this programme they chose to implement directly through NGO partners. According to staff this choice was made deliberately to give WaterAid greater control over the results, predominantly due to the tight timeline (as opposed to due to the PBR modality).

- While Plan International implemented through local government bodies in Bangladesh, this was feasible because programme staff were seconded to these bodies, so as to maintain some control over implementation. Plan International has commented that it would prefer to implement directly through NGO partners in future PBR programmes.

- SWIFT's rural WASH activities were also implemented exclusively through local NGOs in Kenya. In DRC, however, while local NGOs implemented community activities, these had to be scheduled within the framework of the national VEA programme. All NGO staff had to be trained by government staff of the VEA programme and the construction of latrines by households could not take place until the sixth step of the VEA framework was reached (Annual Review 2015). This training requirement delayed SWIFT's implementation. SWIFT therefore opted to implement certain activities in parallel: for example, expediting the distribution of latrine construction kits to target communities (AR 2015).

6.1.7   Effect on relationships with partners and government

PBR is expected to establish a clear and business-like relationship between DFID, the Suppliers and their respective NGO and government partners. **At country level, it is clear that the PBR modality changed the nature of the relationships with local NGO and government partners to some extent.** However, we are not able to comment on the extent to which DFID's relationships with the Suppliers was affected by the PBR modality.

The effect on country-level relationships was most evident in Uganda, which is the only country in the WASH Results Programme where local partners directly involved in implementation were on PBR contracts[18]. There, SNV took the view that the LCBs should undergo the transition from traditional 'partner' to PBR contractor. While relationships between SNV and their LCBs were positive, the contractual relationship strained LCBs' confidence, trust and commitment. In specifying activities in great detail in their contracts, SNV inadvertently sent out a signal that it did not trust its contractors to deliver high-level results in innovative and flexible ways. It exerted a high level of management and supervisory control. While such arrangements increased the likelihood of short-term targets being reached it is still unclear if such arrangements will increase the effectiveness and sustainability of WASH activities.

None of the local partners of SSH4A outside of Uganda, or those under SWIFT and SAWRP, were on PBR contracts (see Section 6.3.1). In general, consortia leads felt strongly that their local

---

[18] SNV also placed local partners in Tanzania on PBR contracts, but these were only involved in monitoring, not in implementation. In Mozambique, local partners supporting government staff were initially on PBR contracts but these were later changed to grant contracts.

partners could not cope with the risk transfer and pre-financing requirements. In Pakistan, SAWRP country staff reported that local partners still felt responsible for meeting agreed targets, even if they were not on PBR contracts – going so far as to contribute their own 'core' funds to cover unexpected expenses. In contrast, SWIFT noted that it was at times challenging to put pressure on local NGOs, especially in DRC, where NGOs were used to receiving grant funding without strict deadlines.

In the case of local government partners, pressure was exerted through regular supervision and monitoring by the country programme managers. Both SNV and SWIFT reported that the pressure to deliver under the programme put a strain on some valuable relationships because government partners were asked to prioritise programme activities over other ongoing projects. In the case of SWIFT in Kenya, for example, relationships became strained when WSUP tried to expedite hardware procurement by NCWSC, BBC Media Action had to put some pressure on local radio stations and one proved unwilling to renew their partnership after the first round of activity, on the grounds that too much was being demanded of them with nothing given in return – they felt that they should be paid for making broadcasts. SAWRP also reported that applying political pressure through Union Council Chairmen or religious leaders was needed to achieve results in some communities where the situation was more difficult.

Overall, the social capital resulting from having established relationships with local NGOs and government partners significantly contributed to maintaining good relationships even under tight timelines. WSUP's established relationship with NCWSC significantly facilitated its work in Nairobi's Dandora slum, whereas Practical Action first needed to set up a memorandum of understanding with NCWSC. Similarly, having national staff that spoke the local dialect and had established relationships with communities facilitated Concern's work in northern Kenya.

6.1.8   Effect on Supplier staff

Most Suppliers felt that the pressure to deliver results and the monitoring and reporting burden associated with PBR was considerable and that it demanded far more staff time than had been anticipated. The pressure to deliver and report under considerable time pressure was felt acutely by staff of the implementing agencies and their local NGO partners. PBR has been referred to by DFID and Suppliers as 'payment by paperwork' and Suppliers' experience was far from the 'light-touch' verification envisaged in PBR literature (see box below). There were three main reasons cited for this pressure on staff: (i) the short period available for implementation of outputs (by December 2015); (ii) the unfamiliarity of staff with the PBR modality (see quote below); and (iii) the fact that the requests for evidence under the verification process were more substantial than expected by Suppliers (see Section 6.2).

> *For SWIFT, the start-up phase took longer than expected as each partner took time to understand ways of working under PBR and the specific interpretation of the payment matrices and M&E and reporting framework […]. All SWIFT partners have had to be more cautious in thinking through the consequences of decisions, which has had a buffering effect on mobilisation, as staff adapt their systems to this operating environment. Preparation, planning and negotiation at county level took a longer time than a more familiar accountable grant mechanism as different stakeholders were not fully conversant in the payment by result methodology, and staff had to be inducted. – AR 2015*

As a result, several country teams found their level of staffing inadequate to deal with the reporting burden which resulted from the verification process. In DRC, SWIFT deployed additional management personnel to strengthen their M&E structure to adhere to the new and more stringent requirements imposed by the MVE supplier (AR 2015). In Pakistan, deploying one engineer and one monitoring, evaluation and research (MER) manager per district was insufficient

to cover tasks such as monitoring data entry, and the other programme staff took turns to help with the task, distracting them from their main work.

In contrast, having the flexibility to re-allocate staff was also seen as a positive benefit of the programme, allowing SNV programme managers, for example, to respond to programme needs.

Box 8: Examples of the 'reporting burden' experienced by Suppliers

- *'The verification cycle is onerous: Every three months we go through verification, meaning there is one month out of every three to implement. The third month [of the quarter] is spent providing evidence, the first month responding to queries from the last round. We don't even have the breathing space to think about sustainability. The most burdened people are the frontline staff. They had to enter data while also delivering the programme. How much does the reporting jeopardise the delivery on the ground?'* – SAWRP (2015 Supplier Learning Event).

- During the initial quarters of the programme verification requirements for SAWRP were still being finalised but implementation was ongoing. As a result, local teams were requested to provide additional information, such as number of under five-years-old children in a household, or the CNIC (identity card number) for households which had already been surveyed. This required local teams to re-visit villages and households to collect the missing data.

- Social organisers and MER staff complained of long working days (often 10–12 hours) and female social organisers highlighted this as being intrusive with regards to their family life, to the extent that they would think carefully before taking on another programme like this.

- There was considerable staff turnover in some of SAWRP's and SWIFT's NGOs (local and international) during the Output Phase. According to staff this is attributed partly to the workload in general, and to onerous monitoring and reporting requirements in particular. One engineer reported that the pressure connected with not hitting a dry well – as this would result in no payment – was so high that this person quit their job. With some key staff having left, there may be problems of institutional memory and continuity as the programme enters the Outcome Phase.

- For SWIFT, the evidence collection process was cumbersome due to the number of people who commented on the evidence prepared for each milestone payment. Before (or during) checks by the Verifiers at national and global level, data would be reviewed, and sometimes commented on, by the partner NGO's management at local, national and (often) international level; by SWIFT management at national and global level, including the MEAL service provider (ODI) on some occasions; and sometimes by DFID. Dealing with this level of scrutiny and feedback was very time consuming for the implementing partners, though the fixed quarterly timetable for reporting and verification was generally adhered to.

- For SWIFT, country managers were more heavily involved in collating and checking monitoring data than had originally been expected. Delivering large volumes of paper-based monitoring reports was logistically difficult for partners working in remote areas.

- SNV also reported that as part of the 'process milestones' it repeatedly had to submit paper-based evidence running to thousands of pages – for example to provide the attendance lists of all the CLTS facilitators that had been trained.

6.1.9   Experience by the verification team

Designing and adjusting the verification approach while implementation activities were ongoing posed some challenges for the MV team (see Section 6.2.1). It is to their credit that the verification approach was sufficiently flexible to accommodate these.

However, both LVs and Country Verifiers felt that the most beneficial part of their role was advising on the Supplier systems needed to accurately track performance and programme risks. Some Country Verifiers felt they did not have enough time to carry out this role well, especially under SSH4A, where the number of days for each Country Verifier was less than in other programmes. The MV team, however, found it challenging to advise Suppliers on their monitoring framework for

sustainability, as – apart from in the case of SSH4A – these were not linked to verification, and thus the MV team had an unclear mandate on the basis of which to make recommendations.

6.2   Scrutiny of the monitoring data

The verification process underlying the PBR modality clearly led to greater scrutiny of the Suppliers' monitoring data. This had several implications for the degree of monitoring effort by Suppliers, as well as the strengthening of the monitoring systems themselves. These are discussed in turn below.

6.2.1   Degree of monitoring effort by Suppliers

Suppliers perceived the reporting burden to be high, compared to what they had originally expected at tender stage, which put considerable pressure on staff (see Section 6.1.8). During the country visits for this evaluation, it was challenging for Supplier staff to quantify how much additional time was spent on monitoring and reporting because these tasks were interwoven with their other activities. One Supplier, however, suggested that the reporting effort negatively affected the time available for programme implementation (see quote in Box 8). While most Supplier staff felt that issues with the verification process did *not* lead to implementation delays – as in many cases evidence requests were made after a portion of implementation work was already completed – they did put considerable pressure on monitoring staff, who had to return to communities to collect additionally requested monitoring data (see box below).

The experience under this programme suggests that the reporting burden was particularly high for process-based milestones because it meant verifying that activities such as training sessions had taken place and the number of people participating, not just checking output (access) data.

The main reason why the reporting burden was perceived to be high was because it was more substantial than Suppliers had originally expected at tender stage. The drivers for this were three-fold:

- **Suppliers did not know how substantial the monitoring requirements would be when they drew up their M&E budgets,** because the verification process and requirements were designed *in parallel* with the Supplier programmes in June to September 2014 (the inception period), rather being completed *before* the Suppliers designed their technical proposals.
- **The limited prior experience of Suppliers with PBR also made it more difficult for Suppliers to know what level of monitoring effort to plan for**. In Kenya, for example, Oxfam had been under the impression at the inception stage that their existing M&E arrangements would suffice, and it was a surprise to be told later by the MV team that significant changes were needed.
- **The verification requirements were modified repeatedly in the initial quarters,** as the full variation of implementation activities became clear to verifiers. While this is an important part of tailoring the verification approach to make it better suited to every country (see Section 5.1.4), the absence of detailed country programme plans prolonged this process as discussions had to be held with every country team to understand the full variation of implementation activities.

SNV staff indicated that as long as the evidence requirements were clear and provided sufficiently in advance, they did not become an additional burden because requirements could be planned for – however, this was not always the case. The box below highlights that for every Supplier there were cases where verification requirements had to be modified or took a long time to be agreed upon, which had consequences for country teams.

> **Box 9: Examples showing the consequences of modifying verification requirements**
>
> - For the verification of Unilever's school hygiene activities under SAWRP, the number of school children attending sessions was requested as evidence. Unilever, however, only provided lists of the number of children enrolled, even though spot checks showed that these numbers were higher than those attending. As evidence on session attendance on the day was impossible to collect retrospectively, payment was partially withheld for several quarters until the issue was resolved: Unilever undertook a survey, concluding that the difference between enrolled and attendance numbers was 5.3%. As a result, withheld payments were released, following a 5.3% reduction.
>
> - SSH4A implemented hygiene promotion through a variety of channels, including door-to-door visits, mass gatherings and radio campaigns, but planned to ensure that households would not be double-counted under several channels. Some of the country programmes decided that it was easier to use radio promotion as the main source of evidence, as this reduced the risk of double-counting when using multiple channels. DFID, however, took the decision that radio-only promotion would not be acceptable, and that any results based on radio promotion would have to be supported by rigorous survey data. As SNV had not conducted any surveys in the Q3-2015 round they were unable to meet this requirement, which led to around 50% of the result not being verified.
>
> - As the SWIFT programme in Kenya involved many diverse partners, tailored verification requirements had to be developed for each one. Several were only finalised in mid-2015, six months before the end of the Output Phase. Partners in Kenya reported that country visits carried out by the LV helped significantly in finalising the verification requirements for each partner.

Based on the Suppliers' experiences, there are clear recommendations for how the verification burden could be lessened:

- At the March 2016 Learning Event, there was consensus among Suppliers, Verifiers and DFID on **the benefits of an inception period for Suppliers**: ensuring that verification requirements are clear before any implementation activities begin was seen as essential, in order for Suppliers to be able to sufficiently resource their support to the verification process. However, SNV cautioned that sufficient flexibility needed to remain to allow verification requirements to be adjusted if necessary, to avoid an overly rigid verification system.

- SAWRP and SSH4A found that the use of a **mobile-based platform for monitoring and reporting** eased the verification process, as each data point included GPS coordinates and could be easily uploaded by field staff, reducing the burden of data entry and data consolidation. However, teams cautioned that mobile-based platforms will require additional training and may not be suited for process-based payment milestones.

- **Undertaking spot checks (small-scale random checks on submitted results) can reduce the scope of the verification fieldwork,** while still encouraging the Supplier to maintain quality across the entire programme (as they do not know where the spot checks will take place).

- SAWRP and SSH4A found that **ground-truthing the monitoring data** – i.e. carrying out their own spot checks before the data were submitted to the verification team – helped to substantially reduce the number of queries which verifiers made on the evidence received. The SSH4A programme manager, for example, carried out a check against the verification forms to make sure the evidence would pass all the verification criteria. In those cases where it did not (e.g. the sanitation survey of South Sudan) this provided the Supplier with the option of retracting their evidence before it was verified, to allow them to complete the gaps.

- The way the governance arrangements were set up between the MV team and the Suppliers meant that there was **no direct communication between the Country Verifier and the Suppliers' country staff.** Communication instead flowed through the LV and the Supplier lead. Suppliers saw this as a key factor in contributing to delays in finalising the verification

requirements. Direct communication would have simplified the process, but it was acknowledged that it could also have caused issues around standardisation of the verification process across countries.

- Suppliers noted the challenge of choosing a realistic **frequency of payment milestones**. Some Suppliers felt that having fewer and more spaced-out payment milestones near programme start would have reduced the reporting burden as there would be more time to collate evidence – especially as behaviour change-based WASH interventions take a while to generate results. This could, however, have caused cash-flow issues. Other Suppliers noted that fewer payment milestones would still involve the same monitoring work as internal monitoring was still done monthly or quarterly.

6.2.2   Strengthening monitoring systems

Despite the burden on staff, Suppliers also reported that the stringent verification requirements resulted in beneficial strengthening of Supplier monitoring systems. This beneficial impact was particularly seen for consortium partners which had weaker M&E systems at programme start. Improvements were seen in the following areas (see box below for examples):

- For most Suppliers, **the reliability of the monitoring data improved**, as a result of more-standardised data collection protocols and more rigorous QA procedures. SWIFT noted that detailed guidance by the country programme manager was key in facilitating this process. More reliable monitoring data provided Suppliers with more confidence in the results they had achieved.
- For some Suppliers, **the frequency of monitoring data collection increased**. This allowed them to be better able to see progress, and adjust implementation based on these trends.
- For many Suppliers, the **more regular interaction between M&E and technical teams** in country, as well as at HQ level, benefitted implementation.

While such improvements are in line with best practice, and can also be achieved through capacity building under a grants programme, it appears that scrutiny of the monitoring systems also helped bring about these improvements. DFID sees leveraging improvements in M&E systems as one of the expected benefits of PBR[19]. Under this programme, however, DFID has not had the opportunity to strengthen government monitoring systems because these were not used by Suppliers.

Verifiers reported that the degree of detail of the verification process made a difference to the *degree of improvements* made to Supplier monitoring systems. Careful attention to detail allowed potential weaknesses to be identified and rectified. Lighter-touch verification may not have had the same beneficial effects on systems strengthening.

Nonetheless, the evaluation has not yet found evidence that strengthening of monitoring systems also resulted in improved quality of implementation. This link will be further explored in 2018 at programme end.

---

[19] Mentioned at March 2016 Learning Event.

Box 10: Examples of improvements in Supplier monitoring systems

- For SSH4A, the Verifier made several recommendations in Q3-2015 on how to strengthen the household surveys, especially after critical issues were better understood following his country visits in mid-2015. As a result, SNV standardised their sampling procedure so that it was undertaken at programme level (rather than at country project level, as previously). The introduction of field spot checks and map checks by the Verifier encouraged SNV to improve its enumerator training, and to ensure that strict protocols were followed for household sampling, and that enumerators were more closely supervised during the surveys.

- For SAWRP, scrutiny of monitoring systems also led partners to put in place better internal QA procedures. As for SNV survey sampling, enumerator training and survey data storage also improved considerably following Verifier feedback on the planned outcome surveys.

- In Uganda, the SNV team put in place a dashboard, where local partners could upload monitoring data through their mobile phones. This allowed programme managers to have a real-time overview of activity progress. Programme managers in the districts met weekly with local partners to discuss progress and find solutions to potential challenges faced in the field.

- In DRC, SWIFT partners set up new monitoring systems following feedback from the Verifier. The new system required weekly progress reports from local partners and allowed the programme to have much better oversight of progress and potential issues arising. This allowed issues to be detected early, as opposed to at quarter-end, and it allowed additional resources to be mobilised, if needed, to reach targets.

- In Kenya, existing monitoring systems of partners were also adjusted following feedback from the Verifier. WSUP, for example, noted that more regular monitoring data was useful to track progress.

- For SWIFT, the Verifier introduced checks of the water quantities delivered per person per day by SWIFT's water installations. Following this, SWIFT also began monitoring whether the water-point facilities of local partners were delivering the per capita per day flow rates set out in their minimum standards. Where water delivery was found to be insufficient to cover population needs, additional natural spring developments were implemented.

6.3   Risk management

Both the PBR modality and the pressure of the MDGs deadline clearly affected how Suppliers managed risk within their consortia. This had several implications for how Suppliers managed their consortia and allocated targets to different partners and countries. These implications are discussed in turn below.

A key challenge for risk management was clarity on the contracting terms: Suppliers felt that a number of programme requirements could have been clearer at the outset, including the eligibility of shared latrines as a basis for payment; and conditions for non-payment, including circumstances under which the force majeure clause would apply. This forms a contrast to one of the expected benefits of PBR: it intends to agree clear and transparent contract terms up front – i.e. knowing exactly what results will be paid for – which should shift institutional incentives for data production, monitoring and programme delivery (DFID 2015a). Clear and transparent contract terms intend to strengthen the mutual accountability of the donor and the Suppliers in regard to reaching the agreed results (Birdsall and Savedoff, 2010) and assume that the donor will withhold funds in the case of non-performance (Clist and Verschoor, 2014).

> *We found that the circumstances under which DFID would take on the financial risk (force majeure clause) was ill defined in the contract between Suppliers and DFID and thereby opened up Suppliers to a lot of risk.* – SAWRP staff

6.3.1   Degree of risk transfer

Under the WASH Results Programme, 100% of payments to Suppliers are made within the PBR framework[20]. Suppliers bear the risk for delivery even in unexpected circumstances. The only exception to the PBR modality is the force majeure clause in the contract, whereby DFID will take on the financial risk under exceptional circumstances. In the event, this applied to the SWIFT Liberia programme, where activities could not continue due to the Ebola crisis. SWIFT was reimbursed for expenditure incurred at inception.

In contrast, experience with PBR in other sectors shows that the value of PBR lies not in reimbursing the direct costs of implementation (as NGOs are already motivated to do this) but in linking PBR payments to additional dimensions, such as quality, sustainability or equity, which might otherwise be marginalised, albeit unintentionally. Evidence from other sectors therefore suggests that there is a strong argument for the bulk of Supplier payments being grant-based, with a relatively small percentage used as an incentive for good performance (NAO 2015). Accordingly, it is common practice in regard to the use of PBR (especially in the UK) to ensure that the share of total payments that are PBR-based remains small; a review of the use of PBR in the UK for example recommends that only 20% of payments be linked to performance (ibid.). This also reduces the burden of pre-financing (see Section 6.3.2). The benefits of a hybrid model were discussed at the March 2016 Learning Event. SNV, however, expressed a preference for a 100% PBR model which focused on quantifiable targets.

PBR theory would have predicted that Suppliers would manage risk by passing it on to their local implementing partners. The Lead Suppliers of SAWRP and SWIFT did pass PBR risks on to their consortium members (all INGOs). However, in many cases, these did not pass PBR risks on to their country offices (e.g. Oxfam GB did not pass risks on to Oxfam Kenya, and SNV did not pass risks on to SNV country offices). Similarly, **the vast majority of Lead Suppliers were not generally willing to put their national implementing partners onto PBR contracts,** on the basis that these partners would not be able to pre-finance their work or bear the risk of not being reimbursed later. One notable exception was SNV Uganda, which decided that it would contract local partners on a PBR basis. In Kenya, the SWIFT implementing partners were in fact all INGOs and Oxfam contracted them on a PBR basis, though local NGOs working under these organisations were funded using grants, as in the past. Under SAWRP, Plan International and WaterAid chose to hold the PBR risk centrally, pre-financing the country teams and other consortium members – which constituted a considerable burden. In practice, payments to consortium partners were always at least a quarter in arrears, making it necessary for them to draw on other internal funds to bridge gaps in their cash flow.

In those cases where the organisations' targets under PBR contracts were not met and payment was withheld, this had serious and far-reaching impacts:

- For Practical Action in Nairobi, their urban water and sanitation project was heavily dependent on action by government partners, and in the end it could not be taken forward to implementation, for a number of reasons. By the time this became clear, the organisation had already incurred considerable costs and therefore suffered a considerable financial loss.
- The PBR contracting arrangement put particular strain on the local LCBs in Uganda: staff interviewed said that, while SNV pre-financed some of their activities, many had to borrow

---

[20] The views of DFID and Suppliers differ on whether this really constitutes a 100% PBR programme. DFID have indicated that the programme is not 100% PBR, because some payments have been allowed for inputs and implementation processes, not just outputs (see Section **Error! Reference source not found.**). However, these ayments were still subject to a full verification process, with the possibility of payments being reduced if some activities were not fully completed and/or supported by evidence. The evaluation team therefore sees the programme as 100% PBR.

additional money to carry out their work. According to interviewed LCBs, some LCBs received a financial penalty for under-performance.

- For the international consortium partners on PBR contracts, some reported using funds from other sources to cover unexpected costs (especially additional staff time) in order to ensure that targets were still met. This was seen as a risk management strategy to reduce the risk of non-payment. Interestingly, however, the same practices were seen among some of SAWRP's local implementing partners. This suggests that the established relationship with local partners instilled a sense of joint responsibility, even without a PBR contract.

6.3.2   Ability to pre-finance activities

PBR is based on the premise that Suppliers are only paid once agreed deliverables have been achieved. This model assumes that Suppliers have funds available to pre-finance large-scale programmes, and that they can spread risk across a portfolio of activities (Clist and Verschoor, 2014).

These requirements played a significant role in determining the membership of consortia: the need to pre-finance activities, and the risk of not being fully reimbursed for work undertaken, posed particular challenges for smaller players with limited reserves. Some bidders interviewed were concerned that smaller NGOs, which might have offered innovative technical approaches, may have been squeezed out by the bigger players. Under the WASH Results Programme, the consortium members needed access to sufficient reserves to pre-finance not only their own start-up activities but also the activities of their local implementing partners who were not appointed on a PBR basis in the majority of cases, unless they had INGO status.

6.3.3   Consortium management

All three Suppliers chose to hold a large portion of risk centrally, pre-financing their other consortium members (SAWRP and SWIFT) or their country teams (for SSH4A). This introduced an additional element of risk for the Lead Supplier, who needed to factor in the likelihood of consortium partners failing to deliver, and setting up procedures for how to deal with such an eventuality.

Both SAWRP and SWIFT explained that managing these partnerships was challenging. Building up the partnership, getting to know each other and the PBR modality and learning to speak the same language took time[21]. Measures were put in place to ring-fence the payments for individual consortia partners, though the legacy of the joint baseline for WaterAid and Plan International meant that for hygiene outcomes the two partners would still be jointly liable.

SNV explained that these challenges of mutual liability were a key reason why SNV chose not to form a consortium, but to implement SSH4A only through SNV's country office and their national partners.

Because the risks were highly centralised, Suppliers managed this risk in three main ways:

1. All Suppliers aimed for beneficiary numbers that were 10%–30% above their contractual target, in the expectation that some results would be disallowed or not achieved. This buffer was designed to ensure that 100% of the target results would be verified and paid for.
2. Suppliers structured their payments over time so that over 70% of total payments were made within the first 18 months of the programme (by December 2015), thereby enabling

---

[21] March 2016 Learning Event.

some early cash flow and reducing the need for pre-financing (see Figure 11). There was consensus at the March 2016 Learning Event that SNV's approach of a high proportion of 'process' payments in the first year of the programme was the most astute in terms of enabling early payments and cash flow.

3. Suppliers incorporated a contingency fund into their budgets (between 4% and 12% of the contract value), which could be available to consortium partners. SAWRP partners reported, however, that these funds were not sufficient to cover all possible scenarios, including the reconstruction of WASH facilities after a major unforeseen event, such as a natural disaster. After 22% of latrines in Umerkot (Sindh province of Pakistan) were destroyed following heavy rains, communities were mobilised to repair the damage at their own expense.

Figure 11:   Comparison of payment timing across Suppliers

The graphs below show that Suppliers frontloaded payments in the Outputs Phase to allow the recovery of overheads sooner during the programme, and to protect against cash-flow issues.

SNV
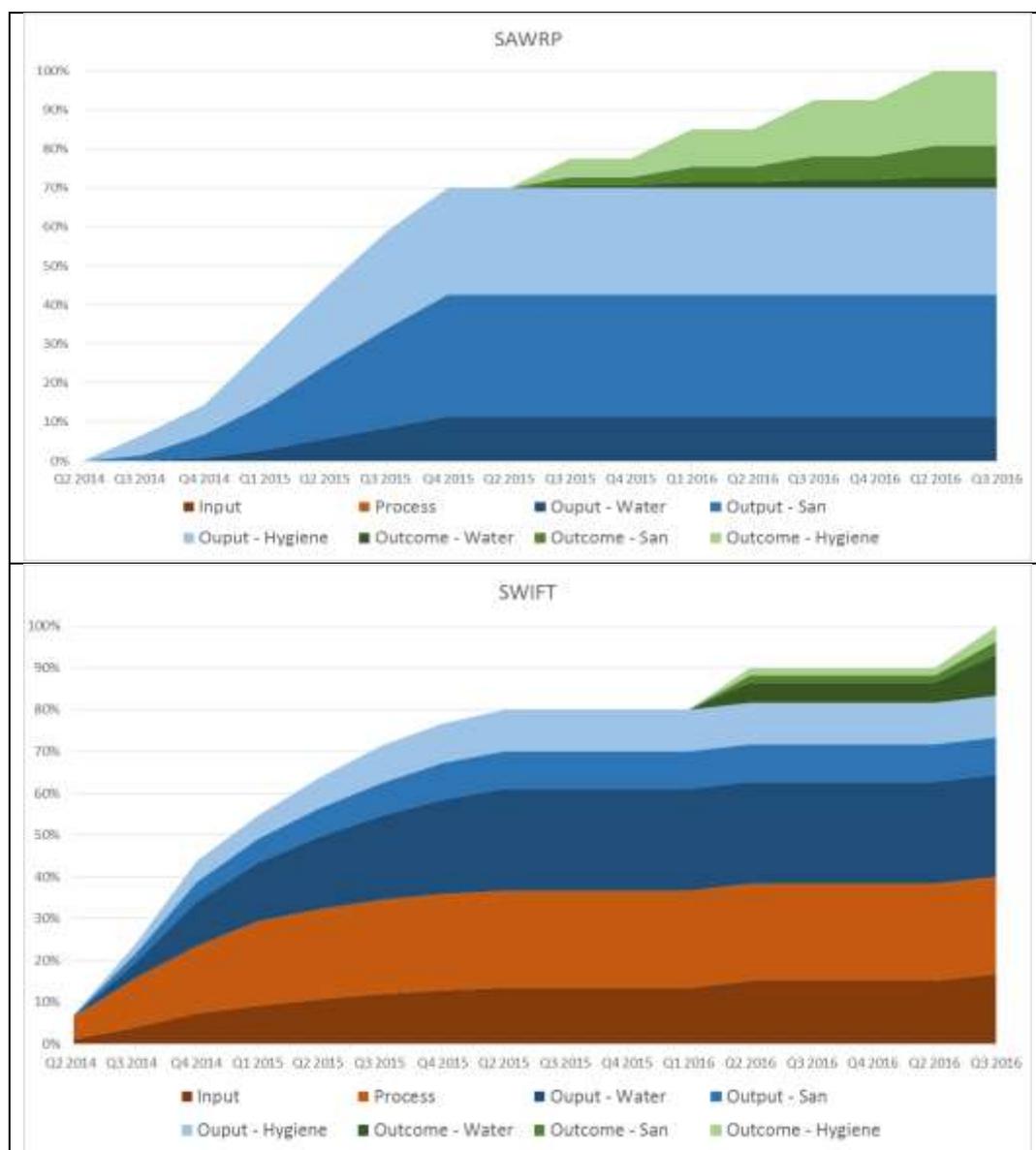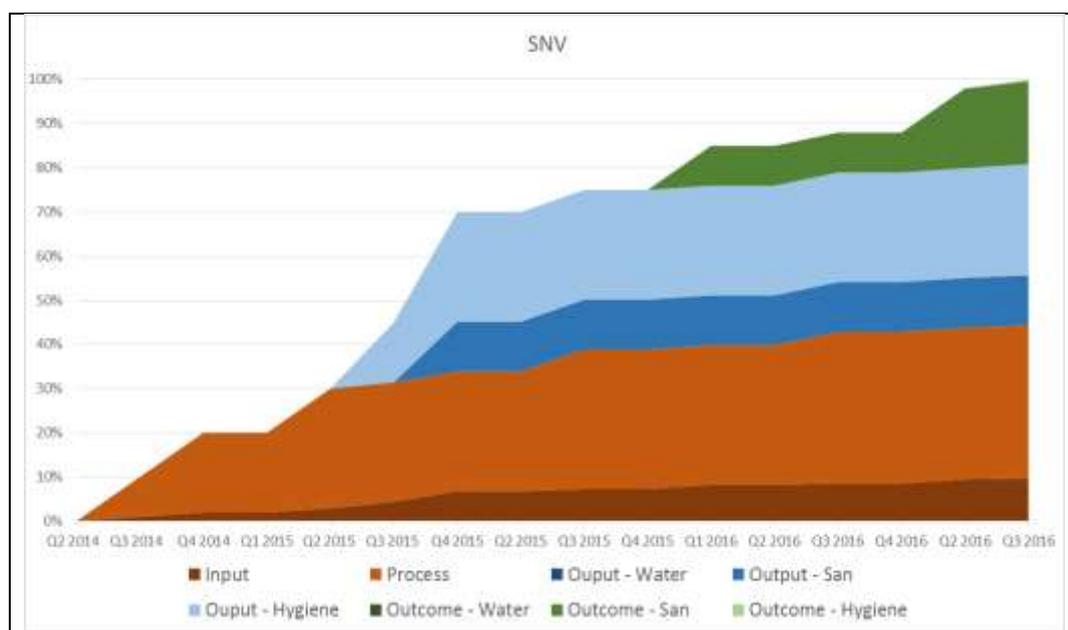
6.3.4   Flexibility in transferring beneficiary targets

A frequently-cited benefit of PBR is the ability to manage risk by transferring beneficiary numbers from high- to low-risk geographical areas or intervention types, if needed. **Partners appreciated the flexibility with regard to transferring results between locations and organisations, but doing so was not always straightforward**, for several reasons:

- unit (per beneficiary) costs varied widely across the sub-projects, and it was essential not to exceed the overall budget limit;

    if taking on extra targets meant moving to a new project area, this was problematic if no baseline survey had been carried out there. DFID approval was needed when transferring results between countries, and this took some time to negotiate; and

- where a memorandum of understanding had been agreed with a district or county government, this limited the partners' ability to pull out of that district and transfer beneficiaries to a different area if there were challenges in implementation.

In the SWIFT programme there were many instances of target results being moved between partners where one was struggling to meet its original commitments while another had the capacity to take on more. These transfers happened both within Kenya and between Kenya and DRC, and in some cases urban results were transferred to rural projects.

In the case of SSH4A, there was only one implementing agency so the option of transferring results between partners did not exist. However, several country programmes exceeded their targets, compensating for others that were unable to meet theirs. In contrast, for SAWRP no transfers were made within or between countries, as all the partners could meet their targets.

The potential for transferring results was clearly helpful for Suppliers, but it also enabled them to focus on (or move to) communities where results could be delivered faster and/or at relatively lower cost, with potential equity implications (see Section 7.4.1).

6.4   Drivers of VFM

6.4.1   VFM in the context of PBR programming

While no quantitative comparison between the VFM of the WASH Results Programme and a comparable non-PBR programme was possible, some qualitative comparisons during case study visits allowed key factors affecting VFM to be explored.

An assessment of VFM under a PBR payment modality attempts to evaluate the extent to which the potential benefits resulting from using PBR are likely to outweigh the costs, relative to other mechanisms (DFID 2014). The benefits can relate to more reliable quantities or improved quality of results, while costs relate to payment levels incurred by DFID. If payment levels are too high, DFID might be able to achieve the same benefits using other contracting mechanisms. On the other hand, if payment levels are too low, this will affect the Supplier's ability to cover their costs, even if VFM appears to have improved.

6.4.2   Determining appropriate payment levels under a PBR contract

PBR theory assumes that suitable payment levels can be identified and agreed that are based on rigorous analysis of the costs involved in delivering activities[22] and in supporting the verification process, including any necessary risk premiums (DFID 2015a). Under the WASH Results Programme, DFID agreed on payment levels in the contract negotiations, which took place under the competitive tendering process.

However, not all Suppliers had a good understanding of how to price in risk in their budget proposals, so that even with 100% achievement of their target results they might not in reality recover all their costs. Suppliers explained that in future they would price in the following costs more accurately:

- the cost of pre-financing partners, namely the risk that partners would not achieve the intended results and that the Lead Suppliers would lose out on the pre-finance payment;
- inflation and salary rises; and
- a buffer for slippage in timing and scope – especially as a result of some M&E activities being required which had not been costed.

The majority of the WASH activities funded by the WASH Results Programme involved incentivising behaviour change by communities, and involved a substantial number of local and government partners. As a result, the Lead Suppliers, who held the risk for non-payment, did not have a very firm control over the results to be achieved[23]. SWIFT proposed a slightly higher price per person than other consortia, to build in the additional cost of working in conflict-affected areas, but other Suppliers did not appear to price in risk in the same manner. There is a forward risk that if these risks had been fully priced in, this may affect the price to be paid by DFID in future. This, however, is equally likely in other non-PBR funding modalities. In summary, VFM under a PBR modality relates less to the actual costs budgeted for, and more to the degree of control of implementers over the achievement of outputs and outcomes (Clist and Dercon, 2014). Overall,

---

[22] Pricing should take into account equity considerations (setting higher prices for vulnerable or lagging groups, such as women or harder to reach areas, DFID 2015a).

[23] The cost of hitting a dry bore-hole, for example, was priced in, and was included in the average price per person of providing water access.

PBR allows DFID to obtain greater certainty regarding the results achieved, thus increasing the 'value' of what is being purchased.

A rough comparison between the price-per-beneficiary paid and the cost-per-beneficiary incurred to date suggests that the price was higher than the cost in most cases. However, discussions with Suppliers showed that this was not necessarily because risk was priced in, but rather because costs which will be incurred later were invoiced for earlier: according to Supplier staff in country, Suppliers have sought payment for the work completed but in some cases have only spent 60%–70% of the invoice – this means they have been able to carry over savings to the Outcome Phase. This provides Suppliers with a buffer to manage risk during the Outcome Phase, where control over results is even less certain than in the Output Phase.

6.4.3   Benefit of flexibility in regard to moving funds between budget lines

Staff interviewed in country explained that the main benefit seen under the WASH Results Programme, in comparison to comparable non-PBR programmes, was having the flexibility to move funds between budget lines. While Suppliers still reported spending against budget lines internally, not having to report and justify to DFID any changes in spending against budget lines gave Suppliers the flexibility to adjust the activities undertaken. For example, spending allocated to one region could be moved to another if increased efforts were needed there, or more funds could be allocated to community visits if increased efforts were needed in this regard. SWIFT stressed that these benefits were particularly noticeable in fragile-state contexts, where it was more challenging to stay within the 10% variation required for each budget line under grant programmes. Nonetheless, Suppliers highlighted that a longer inception phase would have been needed to better familiarise accountants with the new funding modality. However, a fuller VFM analysis is needed to establish whether these benefits outweigh the additional costs which result from the verification process itself, as well as the costs which Suppliers have incurred in relation to supporting the verification process. Suppliers were not able to estimate how much additional time was spent supporting the verification process as this task was interwoven with their other responsibilities. A fuller VFM analysis is planned for late 2016, to better understand these associated costs.

6.5   Learning

The WASH Results Programme has contributed to learning on several levels. First, there has been considerable learning within the Supplier consortia in relation to the demands of a rigorously monitored PBR programme with tight deadlines. This aspect is discussed in detail below. Second, the learning about 'how to do PBR better' has been captured and distilled by the evaluation team and Itad's L&D team, through a number of learning events. Third, there has been some wider sector learning, through the documentation and communication of experiences to date, which will be further considered in the evaluation at programme end. Lastly, the evaluation itself fulfils a learning function in so far as the recommendations that come out of it are intended to allow course correction half way through the programme, if needed.

6.5.1   Degree of learning included at design stage

As the use of PBR in WASH is still relatively novel, DFID stressed that ongoing learning and course correction was intended to be a key part of the programme. To enable learning, a dedicated L&D team was set up within the MV team, with annual Supplier Learning Events organised to

discuss insights arising from the evaluation work and from the After-Action Review process of the verification cycle.

However, Suppliers cautioned that the tight timeline left little room for learning, and that confidentially issues in regard to communication between Suppliers discouraged lesson-sharing between consortia. The absence of a dedicated budget within some Suppliers to contribute to learning initiatives also posed challenges.

To encourage more lesson-sharing and learning, DFID committed to the following in April 2015 (Annual Review 2015):

- *'Invest in learning lessons, and effectively coordinate (all partners). Specifically by sharing the early findings of the programme evaluation, and by partners sharing lesson learning documents where possible.*
- *Better communicate programme findings (DFID, and partners). Specifically by establishing and keeping up-to-date a project website. DFID will initiate summary reports for communications purposes following each verification round.*
- *Increase frequency of lead verifier travel to programme countries (MVE), and in the case of DRC increase country verification capacity.'*

6.5.2   Learning activities undertaken

Across the SSH4A programme, quarterly reflection meetings were held in each country to discuss challenges, and three annual learning workshops were held across all countries. These three workshops focused on different themes: how to design sustainable interventions from the start, how to carry out BCC work and supply chain development, and how do design post-ODF community engagement. Country teams found these to be extremely useful opportunities to exchange comments on experiences and to learn from other countries, especially on how to deal with the tight MDGs deadline and the PBR modality. However, SNV felt that due to the time pressure of the MDGs deadline there was less space for learning than in their previous multi-country programmes.

For SSH4A in Uganda, numerous regular learning events within the programme undoubtedly contributed to improving programme management and increasingly effective pursuit of the defined results. These were mainly conducted through the Chief Executive Officer's (CEO's) reflection meetings and other SNV/LCB experience-sharing workshops. It was, however, difficult in the time available to document processes that could influence wider sector learning. The SNV team observed that documentation is costly in terms of time and money – and is not a priority under the PBR modality. At national level, sharing of best practices is done through the National Sanitation Working Group. SNV has organised learning events for districts, LCBs and Technical Support Unit staff in the project areas to share best practices and further develop strategies for performance improvement.

For the SAWRP consortium, the learning partner (WEDC) facilitated inception workshops in Bangladesh and Pakistan in May/June 2014, a learning event in each country in February 2015, and an outcomes workshop in each country in November 2015. Programme staff in Pakistan felt that the Output Phase had been a great learning experience in terms of how to deliver high quality results at scale within a short timeframe. However, given the urgency of the MDGs deadline there was resistance from district teams to participating in learning exercises because of time constraints and the possible effect on meeting their targets. The Lead Supplier commented that due to the PBR modality there was less clarity at management level regarding which activities were being

implemented on the ground, but more clarity on which results were being achieved. This, however, made it more difficult for WEDC to guide learning on implementation experiences.

For the SWIFT consortium, the learning partner (ODI) facilitated various workshops on implementing WASH in fragile contexts and on how to design monitoring frameworks. A dedicated SWIFT website was established,[24] where short case studies and briefing notes on DRC and Kenya were produced. Programme staff in Kenya felt that while there were regular meetings between partners to share information on progress and working under the PBR modality, structured learning within the Kenya country programme has been quite limited to date. The urgency of the MDGs deadline meant that partners had been too busy to give much attention to learning. To support country teams on their learning, SWIFT commissioned two independent reviews during the Output Phase: one focusing on programme experience with PBR and the other being an internal mid-term review. The latter was not made available to the evaluation team at the time of writing.

Finally, the L&D group within the MV team also invested in learning. In addition to the Annual Learning Events organised for the whole programme, the MV team also held annual internal workshops with the evaluation team to consolidate the lessons learned about verification. It was, however, challenging to find enough time for reflection given the urgency of the MDGs deadline. More resources would also have been needed to better draw on the learning provided by Country Verifiers. To better separate the learning role from the verification role, it was discussed at the March 2016 Learning Event that it may be beneficial to situate the L&D group within the evaluation team in future programmes.

---

[24] See www.swiftconsortium.org.

7   Forward view

This chapter looks ahead as the programme enters the Outcome Phase. It reviews the Outcome Phase strategy, targets and monitoring arrangements, and examines the prospects for delivering equitable and sustainable benefits and, ultimately, achievement of the intended programme impact.

Key findings of this chapter
• For two of the Suppliers, outcome targets were only finalised at the end of the Output Phase, with DFID providing some direction on minimum levels of attainment. For water and sanitation, outcome targets were set at 70%–100% of the output target levels. For the adoption of hand-washing with soap, however, outcome targets were lower, at 10%–15% of the output target levels.
• Outcome Phase strategies have been agreed, but detailed monitoring framework and verification protocols are still under discussion. For most country programmes, the start of the Outcome Phase signals a shift in programme focus from direct implementation to capacity building and working with government to ensure that benefits can be consolidated. However, in some countries Output Phase activities will partially continue.
• Outcome Phase results will be measured mostly via surveys; in addition, Suppliers and Verifiers will monitor and track whether enabling conditions for sustainability are in place. Putting in place the preconditions for sustainability will be a priority during the Outcome Phase.
• On the prospects for equity, several examples are noted of cases where Supplier activities were deliberately designed to reach potentially excluded groups. Suppliers have also put in place mechanisms to monitor whether poor people are in fact being reached by the programme, though monitoring could be strengthened.
• While there are some concerns relating to the likelihood of health impacts, the degree to which these are achieved will only become apparent some years from now.

7.1   Outcome Phase plans and targets

For Suppliers, setting the level of ambition for programme outcomes has not been straightforward given that these relate to matters beyond the Suppliers' direct control – particularly hand-washing practices. After much debate during the Output Phase, DFID took advice from the MV team on what level of achievement (i.e. outcomes as a percentage of output numbers) could reasonably be expected for the three outcome targets, based on sector experience. In some cases, DFID considered that Suppliers were not being ambitious enough. On the other hand, SSH4A's sanitation outcome targets were more ambitious than other Suppliers. For example, DFID indicated that it was reasonable to expect that 75% of the beneficiaries at the output stage would continue to use toilets by programme end (one Supplier had proposed a lower target), but for hygiene DFID accepted that 10%–15% of beneficiaries continuing to practice hand-washing with soap was reasonable, based on the MV's review of the literature. The final outcome targets agreed with Suppliers are set out in Table 13 below.

Table 13:   Outcome Phase targets by Supplier

Outcome targets	SAWRP	SWIFT	SSH4A
Sanitation	75% of poor people reached between April 2014 and December 2015 continue to use basic or improved project latrines by the end of March 2018.	70% of the beneficiaries that are reached at output level continue to use latrines that meet SWIFT standards.	100% of output beneficiaries access and use improved sanitation (DFID definition). 31% of Output Phase beneficiaries upgrade to improved sanitation (JMP definition).
Hygiene	10% of poor men, women and children reached with hygiene promotion during the Output Phase continue to practise hand-washing with soap at critical times.	15% of the beneficiaries that are reached at output level continue practising hand-washing with soap at critical times.	15% of the beneficiaries that are reached at output level continue practising hand-washing with soap at critical times.
Water supply	90% of poor people reached between April 2014 and December 2015 continue to use reliable, safe drinking project water sources by the end of March 2018.	75% of the beneficiaries that are reached at output level continue to use SWIFT water-points / water systems.	*Not applicable (no water supply component).*

For most implementing partners, detailed Outcome Phase workplans were still being developed at the time of the country visits, though by September 2016 the majority have been signed off. **For most country programmes the start of Outcome Phase[25] signals a shift in programme focus and some reduction and reorganisation of programme staffing. However, in some countries Output Phase activities will partially continue.** In the case of SWIFT, for example, rural sanitation promotion in Kenya will continue to focus on achieving ODF status and WSUP will continue pursuing 3,000 new water connections in Dandora, Nairobi.

Apart from an absolute reduction in numbers, staff turnover has been a challenge for some sub-projects and there is anecdotal evidence that onerous reporting requirements coupled with rigid deadlines made the Output Phase very stressful for some field staff and managers. Some projects may now face problems of continuity and institutional memory where key staff have left.

A common feature of Outcome Phase plans across the three Suppliers is an emphasis on fostering a greater role for local government agencies in supporting and enabling the sustainability of ODF status and/or the long-term functioning of water supply services. For example:

• for SWIFT's urban water supply projects in Turkana, Oxfam support will switch from infrastructure development to institutional support to improve service delivery;

• SAWRP Outcome Phase objectives are to secure the continued use and maintenance of sanitation and water supply facilities installed in the Output Phase, and the continued practice of hand-washing with soap at critical times; and to strengthen sector management and coordination by government at local, regional and national level, enabling effective approaches applied by the programme to be continued and applied more widely in support of national goals. For schools, the focus of attention will shift from direct implementation to securing

---

[25] SSH4A has a unique phasing structure which is not based on the simple Output / Outcome Phase separation used by DFID at the global level. In Uganda, the fourth phase of SSH4A corresponds closely to the start of the Outcome Phase for the WASH Results Programme overall.

ongoing hand-washing promotion within the curriculum. Plan International, while continuing to engage directly with target communities to consolidate latrine use and hygiene behaviour change, will also second some its key regional staff to district government offices to focus on local government capacity building. There is, however, some concern regarding the low level of interest from Union Council secretaries and the limited budgets available to them; and

- for SSH4A, its results framework and workplans included measures from the start that would lay the foundations for sustainable outcomes, so the latter part of the programme does not represent a distinct shift in focus. In Uganda, the clearest written expression of SNV's intentions is set out in its Phase IV contracts with its implementing partners, local NGOs (known as LCBs). These state that the LCBs should *'work with the existing local government structures to plan and implement interventions towards sustainable sanitation and hygiene promotion in the above districts … ensure that the villages that claim ODF status are linked to the parish and sub-county structures, where the latter will continue to monitor and oversee sustenance of ODF and improved hygiene and sanitation behavioural practices in the respective villages. The LCB will also be expected to facilitate learning and quality assurance between the villages….'* It is clear that SNV's intentions in 2016 have been appropriately and rightly focused on institutional and community capacity development, inclusion at village and household level, behaviour change and sustainability.

7.2   Outcome Phase monitoring by Suppliers and the MV team

7.2.1   MV progress on outcomes by Suppliers

In the Outcome Phase, verification of outcomes will be exclusively based on the household surveys undertaken by Suppliers. By mid-2016, the Suppliers had all developed their MV approach for the Outcome Phase. While SSH4A had most of the details in place from the start, for SAWRP and SWIFT the plans emerged gradually over the course of the Output Phase and involved lengthy negotiations with the MV team and DFID.

SAWRP has drawn up a Monitoring and Verification of Outcomes framework. This sets out that SAWRP will conduct three annual outcome surveys each for water and sanitation outcome indicators. On hygienic hand-washing, behaviour change will also be assessed twice during the Outcome Phase, both for Unilever's work in schools and twice for Plan International and WaterAid's work in rural communities. The outcome surveys will be undertaken by project staff, with technical support and advice from Ipsos MORI and WEDC. Field staff will be trained on the survey approach, and it is proposed that the assessors will work outside their Output Phase areas of intervention. This will also encourage peer to peer learning.

For SWIFT, progress on outcomes were measured through a mid-line survey in September 2016, followed by an endline survey in September 2017. The Sanergy and BBC Media Action projects will only be surveyed in September 2016, not at programme end, because a two-year gap was considered too long in regard to measuring recall of hygiene messages. The survey methodology and design will be led by ODI.

In the case of SSH4A, SNV will continue to track progress on sanitation and hygiene behaviour change through annual household surveys in Q4-2016 and Q4-2017, quality-assured by the MV team. The household surveys will be undertaken by either project staff, government counterparts or SNV consultants, depending on the country.

7.2.2   Monitoring of sustainability prerequisites by Suppliers

SSH4A was the only programme where the Supplier had its own sustainability strategy and monitoring framework in place from the start. SAWRP and SWIFT developed sustainability frameworks over the course of 2015, and both programmes have now adopted a sustainability strategy and monitoring framework.

Under SAWRP, the consortium partners have adopted Sustainability Assessment Frameworks (SAFs), developed in collaboration with WEDC and Ipsos MORI. The SAFs focus on environmental sustainability, institutional/financial sustainability, and equity/inclusion.

These SAF assessments will be administered annually over 2017 and 2018, using teams comprising staff from partner national organisations or national consultants, supported by WEDC. A RAG analysis will be used to assess progress and performance against key sustainability objectives. It should be noted that the two main consortium partners have had to commit some of their own funds to cover the costs associated with development and implementation of the SAFs.

SWIFT has shared a draft sustainability monitoring framework with the MV team. Sustainability will also be monitored annually, with different functional, institutional, financial, environmental sustainably and equity indicators proposed for every NGO partner.

For SSH4A, sustainability aspects will be monitored based on 10 indicators (see box below) which track the existence of various technical, environmental and institutional prerequisites. For some of these, indicators will be monitored through the household surveys (on these, monitoring already began during the Output Phase). SNV is still finalising its methodology for how the remaining sustainability aspects will be monitored. A consolidated monitoring report on all 10 sustainability aspects will be produced twice, in Q2-2017 and in Q1-2018.

Overall, monitoring the enabling conditions for sustainability both by the Suppliers and the MV team provides a strong framework to track progress and flag potential risks to sustainability. As the details of the sustainability indicators and measurement methodologies are still to be agreed with DFID, these will be evaluated in more detail in 2018.

Box 11: SNV's SSH4A 10-point sustainability framework

Strengthening capacity for steering and implementation of sanitation demand creation, with the objective that local organisations are capable of implementing and steering sanitation demand creation at scale

Indicators

S1. Capacity of local governments or line agencies to steer sanitation demand creation at scale in their area

S2. Capacity of local organisations to implement sanitation demand creation (CLTS) with quality

Strengthening capacity for sanitation supply chains and finance, with the objective that affordable market-based solutions for a variety of sanitation consumer needs are implemented at scale

Indicators

S3. Progress on private sector engaging in sales of sanitation hardware and services to bottom of the pyramid

S4. Availability of affordable sanitation options for the poorest wealth quintile. This indicator will compare the cost of sanitation options with the income in the lowest wealth quintile. Affordability should not exceed 5% of annual cash income. Income data are collected in a household survey or from secondary sources, if these exist. The cost of sanitation options will be based on the inventory held by the private sector

Strengthening capacity for BCC for hygiene promotion, with the objective of anchoring effective hygiene BCC in local practice

S5. Progress on institutionalising hygiene BCC. Measuring the existence and quality of a BCC strategy related to hand-washing with soap

Strengthening capacity for WASH governance, with objective of improving local WASH governance terms of alignment of stakeholders, sector planning and monitoring, transparency and social inclusion

Indicators

S6: Improved sector alignment at local level

S7: Progress on the influence of women in rural sanitation and hygiene programmes

S8: progress on the influence of poor households and minority groups in rural sanitation and hygiene programmes

S9: progress on the influence of disabled people and elderly in rural sanitation and hygiene programmes

Improving environmental management through improved faecal sludge management

Indicator

S10: Progress in safe management of faecal sludge

7.3   Prospects for hygiene behaviour change

Regarding how effective the outputs achieved will be in bringing about sustainable behaviour change, this section discusses the prospects for hygiene behaviour change specifically. To date, the programme has only tracked the delivery of hygiene messages: it has not yet tracked the effect of these messages on hand-washing behaviour. There is little consensus in the WASH profession as to what constitutes an adequate hygiene promotion intervention – i.e. one which gives reasonable likelihood of translation of programme activities into changed behaviours. DFID's protocol on counting numbers of people 'reached' by hygiene messages does not include a detailed definition of what constitutes adequate hygiene promotion. A review of sector experience by the MV team found that even 'good' hygiene promotion interventions tend to only result in 10%–15% of those reached by hygiene messaging changing their hand-washing behaviour.

It is too early to assess the likelihood that hygiene messaging implemented in the Output Phase will change hand-washing behaviour in the coming two years. SSH4A results, however, seem promising: in addition to tracking delivery of hygiene messages, SSH4A also tracked the presence of hand-washing facilities in households – this has improved on average by 25% over the baseline (see Table 14). However, hand-washing behaviour change has not yet been reported on.

Table 14: **Progress on presence of hand-washing facilities for SSH4A countries**

	Percentage of households with presence of hand-washing facility (with or without soap)		
	Baseline	**Second mid-line**	**%change**
Ethiopia	2%	33%	31%
Ghana	<1%	6.5%	6%
Kenya	5%	8.5%	4%
Mozambique	27%	37%	10%
Nepal	17%	48%	31%
Tanzania	<1%	42%	42%
Uganda	2%	23%	21%
Zambia	1%	25%	24%
South Sudan	n/a		

7.4 Prospects for equity, sustainability and health impacts

7.4.1 Prospects for equity

While equity was not explicitly built into the PBR framework – payment triggers relate only to the total number of beneficiaries delivered, irrespective of their status in terms of wealth, vulnerability, gender and disability – we highlighted in Section 4.8.4 that all Suppliers planned to implement activities in poorer areas with low access to WASH services. Looking back over the Output Phase, it is evident that the majority (if not all) of the communities targeted by the programme are overwhelmingly poor. Several examples are noted of cases where Supplier activities were deliberately designed to reach potentially excluded groups (see box below).

Suppliers have also put in place mechanisms to monitor whether poor people are in fact being reached by the programme. SSH4A's surveys disaggregated the beneficiaries reached by gender and wealth quintiles during the Output Phase and found that a significant portion of beneficiaries were in the two lowest wealth quintiles. SWIFT indirectly monitored how benefits were distributed within targeted communities by putting in place a beneficiary feedback system: Oxfam's partners conducted community FGDs regarding accountability every quarter, and put in place a complaints mechanism; Tearfund has also put in place an anonymous complaints mechanism (AR 2015).

However, there was a risk that PBR would discourage serving the poorest and most difficult-to-reach communities or locations with low population density, due to the requirement to meet targets at an acceptable unit cost. *A priori*, this risk exists for all funding mechanisms without an explicit requirement to target the poorest, not just for PBR programmes. However, the flexibility of the PBR framework – allowing Suppliers to transfer some or all of their target beneficiaries from one partner to another and from one location to another – compounded this risk, as it gave Suppliers the option to transfer beneficiaries – this is usually not an option in grant programmes. The evaluation documented several examples for all three Suppliers of cases where beneficiaries were transferred between geographic areas. None of these examples are known to be cases indicating a shift to less poor communities, though one example was seen of a shift to an area of lower operational costs[26]. Nonetheless, it raises the question of what is meant by

---

[26] In northern Kenya, PBR-related risks encouraged partners to work in communities where beneficiary numbers could be achieved with the least operational challenges, and hence lower cost. This led to Oxfam focusing its greatest efforts

'results', and whether it is only the number of beneficiaries that matters. DFID mitigated the equity risk by requiring DFID permission be given before any beneficiary targets could be transferred between countries.

It will be important in the Outcome Phase for all Suppliers to monitor equity aspects explicitly so that operational approaches can be adjusted where necessary to ensure a truly inclusive approach. Programme monitoring might, for example, reveal the need to modify water tariff structures to ensure affordability for the poorest households and/or to ensure that cheaper designs, materials and skilled labour are locally available to enable all households (including disabled people) to access (or upgrade to) improved latrines.

Box 12: Examples of activities designed to reach potentially excluded groups

- As the SAWRP programme had an explicit equity focus, it concentrated on geographic areas which had low WASH coverage at baseline. However, no explicit strategies were pursued to reach under-served pockets within areas of higher coverage.

- In Kenya, BBCMA undertook community outreach work to obtain feedback from communities. These discussions showed that women have lower radio ownership and are less likely to be reached by hygiene messaging. BBCMA are now considering distributing the previous years' recorded radio programmes, saved on a USB stick, to allow for a wider listenership.

- SWIFT's water supply interventions paid special attention to the needs of the poorest households – for example by establishing shared connections and/or ensuring that tariff levels were affordable for the poorest households. In Lodwar, urban water supply improvements also included the development or improvement of kiosks where people without access to a house connection could purchase water at very low cost.

- In Ethiopia, SNV's activities ensured inclusion of the different categories of vulnerable people, such as the elderly, persons with disability and widowed/female-headed households, by ensuring that the information materials were inclusive. Households that lack access to sanitation facilities were supported through the social support services.

7.4.2   Prospects for sustainability

Prospects for sustainability refers to the duration of programme benefits after the programme has ended, not just up to the end of the Outcome Phase, during which the Suppliers will remain active in the target communities. Sections 6.1.5 and 6.1.6 set out how Suppliers succeeded to varying degrees to put in place prerequisites for sustainability.

Sustainability has been built into the PBR framework in so much as Suppliers have been incentivised to ensure that an agreed level of results is sustained at least until programme end. Suppliers report that this has encouraged them to consider more carefully how achieved results can be sustained at least until programme end. It has also encouraged discussion on what levels of slippage are realistic based on sector experience. The MV's role in tracking risks to sustainability and equity has arguably played a role in bringing these risks to the attention of Suppliers.

In addition, SSH4A explicitly linked certain sustainability prerequisites to its payment milestones.

At this point, the best prospects for sustainability arguably lie with the urban water supply projects, whereby poor households (or their landlords) have become paying customers of a regulated water

---

in two areas (Kakuma and Lodwar) though water supply projects were spread over a much wider area and included numerous remote rural locations. Despite these influences on village selection, the communities served were undoubtedly very poor.

utility – especially given that these utilities are (or soon will be) receiving technical assistance and capacity building support from the programme and/or other external agencies.

For rural WASH, some projects have benefitted from the use of high quality, durable water supply hardware and hybrid systems incorporating solar pumping to reduce the risk of power failure. This is commendable. Nevertheless, the sustainability risks are greater than for urban projects given the dependence (in most cases) on informal, community-based management systems. Sector experience shows that community management requires a supportive local environment, in terms of technical and motivational support from local government agencies and the ready availability of essential spare parts. As highlighted earlier, a key concern is the capacity and motivation of government agencies to fulfil this supportive role to secure the long-term functionality of water supply schemes; the maintenance of ODF status and the emptying, maintenance and (where necessary) upgrading of household toilets; and the practice of key hygiene behaviours. These risks would apply whether this was a grant-funded or a PBR programme.

7.4.3   Prospects for health impacts

Section 4.8.5 explains that DFID's design did not set any requirement regarding the density of WASH interventions, or the degree of integration of water, sanitation and hygiene activities. It highlighted the concerns for how these design features may affect the prospects for longer-term health benefits. Whether the programme delivers health benefits in the long term remains to be seen.

The degree of integration of water, sanitation and hygiene activities was not unusually low. The extent of focus on geographical saturation, however, varied: while some Supplier projects have indeed pursued a geographical saturation approach to sanitation and hygiene promotion, others have not done so.

Looking forward towards the Outcome Phase, some concerns regarding the prospects for longer-term health benefits therefore remain. The prospects for equity, sustainability and health impact will be explored further when the three case study countries are revisited in 2018.

8   Conclusions

DFID's WASH Results Programme has several unusual or important features, the most significant of which are:

- its scale (approximately £75 million over four years, to serve 4.5 million unique individuals);
- its contractual / financing modality (PBR), in turn necessitating a strong focus on rigorous and verifiable monitoring;
- the 'hard deadline' of December 2015 for delivery of beneficiary numbers ('outputs');
- its 'shape' (the Output and Outcome Phases);
- the expectation that services would reach the poorest and least advantaged; and
- the expectation that large numbers of individuals would continue to use improved sanitation and water services and to practise good hygiene, beyond the life of the programme; and that this would lead to reduced diarrhoea and child mortality.

The evaluation of the programme offers a number of opportunities for DFID, the programme Suppliers and the wider WASH and international development sectors to learn from the experiences to date. The evaluation so far has comprised several components: design of an evaluation methodology centred around the theory of change; detailed analysis of programme documentation and the perspectives of key informants within DFID, the MV and Supplier staff at HQ level; undertaking three country case studies, enabling in-depth discussions and observations on the ground; and the design and implementation of an RCT in one country. This evaluation report summarises the above – except for the findings of the RCT, which are presented in a separate report.

Chapters 5, 6 and 7 have presented the findings of the PEW activities to date. This chapter offers an opportunity to step back and reflect on the overarching evaluation questions, and the key conclusions emerging from the evaluation to date.

8.1   Evaluation questions answered

Based on the evidence and analysis presented in the preceding chapters, the table below summarises our current answers to the evaluation questions set out in the Evaluation Design Document, and outlined in Section 3.4 above.

Table 15:   Evaluation questions answered

DEQs	Our current conclusion	Report section	Confidence level
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?	Type and scale of anticipated results were clearly defined for outputs, but level of ambition at outcome level was much less clear at the design stage. Objectives relating to equity and sustainability were stated but their relationship to the PBR modality was not explicit in the Output Phase.	4.1, 4.2, 4.3	HIGH
DEQ 1.2: To what extent does the **programme's design** (i.e. the theory of change) set out a clear and realistic process for how programme activities will achieve intended outputs, outcomes and impacts?	The theory of change lacked important detail: assumptions were not clearly set out, and the rationale for, and contribution of, PBR was not explained.	4.4	HIGH
DEQ 1.3: To what extent were the **scale and pace of the programme** (including the Dec. 2015 deadline) realistic for achieving the intended outputs and outcomes given the capacity of Suppliers and their local partners?	Suppliers explained how/why they set realistic targets and selected partners. But the urgency of the Dec. 2015 deadline in some cases affected quality and sustainability of results.	4.6	HIGH
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?	Rationale for using PBR unclear at start, but DFID has clarified that priorities were to stimulate the formation of new partnerships for delivery at scale and to achieve greater certainty around the delivery of results (quantity and timing). These were achieved, but PBR made Suppliers risk-averse in setting targets. Incentives were framed negatively in original design: penalties for not meeting targets but no bonuses for over-delivery. PBR modality not wholly appropriate as equity / sustainability aspects were not built into payment criteria. No official inception phase for Suppliers but considerable inception work needed to design MV systems given that programme design was new for DFID and Suppliers. Government engagement is important for sustainability but it was challenging to meet output deadlines via government agencies that are not subject to PBR.	4.7 5.1	HIGH Minor risk of obtaining different opinions within DFID
DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?	Some private sector bidders were put off by uncertainty around how the PBR modality would operate.	4.8.1	HIGH

DEQ 1.6: How likely was it that the programme design would encourage Suppliers to propose 'innovative' WASH interventions"?	Not likely; PBR risks and tight timeline discouraged innovation and encouraged Suppliers to adopt (mostly) familiar, tested approaches. However, programme management was flexible and responsive to changes on the ground, with many examples of results-oriented problem-solving.	4.8.1	
DEQ 1.7: How likely was it that the programme design would encourage inclusive outputs and outcomes?	Unclear at this stage; Suppliers have targeted under-served locations but data unavailable on extent to which results so far have been inclusive and benefitted the lowest wealth quintile. DFID did not specify a targeting process or enforce the need to report on disaggregated results.	4.8.3	**HIGH** Minor risk of obtaining different opinions within DFID
DEQ 1.8: How appropriate was the WASH Results Programme's design for achieving the programme 'learning objectives'?	Somewhat appropriate; there has been some learning at global level and within Supplier organisations, but tight timeline left insufficient space for this, and no explicit lesson-sharing mechanism was built in at the start.	6.5	
DEQ 1.9. To what extent was the **design of each consortium sub-programme** appropriate for achieving DFID's key objectives?	Designs were appropriate insofar as they enabled delivery of the scale and type of results sought by DFID. However, equity, sustainability and government ownership were not built into payment criteria and in at least some projects these aspects were marginalised during the Output Phase.	4.5	

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?	Each Supplier has achieved its total Output targets, but few sub-projects target entire districts/sub-districts or operate via the government framework. 'At scale' was not defined in the programme design.	**6.1.1**	**HIGH**
DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?	A critical challenge for all Suppliers was the tight MDGs deadline; nevertheless, targets were met. Complexity of consortium membership was also key. To be further explored in 2018.	**6.1**	**MEDIUM** Contingent on interviewees having a detailed knowledge of driving factors across different countries. Contingent on observing different levels of achievements across countries
DEQ 2.5: Under which circumstances did the **PBR framework** help/hinder the achievement of intended outputs and outcomes?	PBR risks sharpened minds on results and each Supplier over-delivered outputs. However, the need for cash flow also led Suppliers to include payment for inputs in their results frameworks, which added considerably to the reporting and verification burden.	6.1	**MEDIUM** Contingent on interviewees being able to compare

DEQ 2.6: Under which circumstances did the **PBR framework** affect the quality of programme implementation (positive or negative)?	The quality of processes and outputs was generally good, but in a few isolated cases, the MDGs deadline and associated payment risks led implementing agencies to adopt a 'light' version of their operational model or waive established procurement rules.	6.1, 6.2	experiences to a relevant non-PBR programme
DEQ 2.7: Under which circumstances did Suppliers implement innovative approaches and focus on learning?	Little evidence of Suppliers adopting innovative operational approaches, though management was flexible and responsive to local circumstances. On learning, see 1.8 above.	6.1.4	**HIGH**

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?	Many aspects of the global programme design were unclear at the bidding stage. This resulted in a long procurement process and inception period, which – in addition to the long design phase – reduced the time available for implementation.	2.2.3	**HIGH** Minor risk of not achieving a balanced view of the tendering process between bidders and DFID Minor risk of not interviewing key stakeholders who have left (institutional memory)
DEQ 3.2: Under which circumstances did the **PBR modality** affect cost-effectiveness of individual sub-programmes?	PBR allowed flexible use of budgets but also created cash-flow challenges, as Suppliers had to pre-finance activities. The reporting and verification burden was considerable, creating significant additional costs. Data unavailable on whether results were delivered with greater efficiency or VFM than would be the case with other financing modalities.	6.4	**HIGH**
DEQ 3.3: Under which circumstances did the **PBR modality** strengthen programme monitoring and management arrangements of individual sub-programmes?	Suppliers had to strengthen their monitoring and reporting systems to meet verification requirements. This provided DFID with greater confidence regarding the results reported. However, it is still uncertain at this stage whether the onerous monitoring and reporting burden added value to the quality of programme implementation.	6.2.2	
DEQ 3.4: Under what circumstances did consortium complexity affect the efficiency of **programme management** arrangements of individual sub-programmes?	Programme management burden increased with the number of partners and range of programme activities.	6.3	**HIGH** Minor risk of not achieving a balanced view from programme implementers

HEQ4 Impact: How likely is it that the programme will achieve its final impact objectives while minimising unintended negative consequences?			

			MEDIUM
DEQ 4.1: How likely is it that the programme will achieve its health and non-health **impacts**?	Uncertain; will consider further in 2018, but notable that DFID did not require the achievement of ODF status or the concentration of water, sanitation and hygiene interventions in the same locations.	**7.4.3**	Challenging to assess validity of evidence across highly varied country experiences
DEQ 4.2: Under which circumstances did WASH Results Programme activities have any **unintended/ unplanned positive or negative impacts**?	Uncertain; will consider further in 2018. Considerable staff turnover in some projects and anecdotal evidence that in some cases this was the result (at least in part) of stress related to tight deadlines and onerous reporting requirements. Staff turnover has negative implications for programme continuity and institutional memory.	**6.1.8**	MEDIUM

Contribution of the WASH Results Programme to unintended impacts is harder to assess at community level

Challenging to assess validity of evidence across highly varied country experiences |
| **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? | Risks are evident: groundwork for sustainability marginalised in some projects due to tight output deadline; having separate Output / Outcome phases may also have encouraged this inadvertently.

To be further explored in 2018. | **7.4.2** | MEDIUM

Challenging to assess validity of evidence across highly varied country experiences |

9   Implications and recommendations

9.1   For the Outcome Phase

There has been tremendous effort and goodwill on all sides to make the WASH programme a success, and Suppliers have delivered on their output commitments. The recommendations which follow therefore continue in the same spirit, by identifying key focus areas to ensure that the Outcome Phase results are delivered as fully as possible.

Ensuring effectiveness

1. **Ensuring equity and inclusion:** DFID should make clear to Suppliers and the MV team the importance of aiming for all members of communities targeted in the Output Phase to have access to water and sanitation services and hygiene promotion efforts. In the spirit of the newly-entered Sustainable Development Goals era, the pledge that '… *no one will be left behind*'[27] should apply. To the extent that this has not yet been addressed in the Output Phase, it should be a significant focus of the period to 2018. Specifically, Suppliers should disaggregate their outcome results by gender and wealth quintile, to allow equity to be monitored, and should specify which implementation approaches will be used to ensure potentially excluded groups are reached. This will increase the likelihood of the intended health impact of this programme being achieved, by encouraging a higher geographic concentration of interventions.

2. **Capitalising on learning:** Given the novelty of PBR in the WASH sector it is important that further learning is prioritised and documented, to allow the programme to contribute to strengthening capacity in the WASH sector. In particular, we recommend that:

    a. **a guidance note on how to measure the adequacy of hygiene promotion activities and the extent of hygiene behaviour change** is produced, drawing on lessons from programme experience. Better understanding the drivers of hygiene behaviour change under different contexts could also be an explicit focus of the endline evaluation in 2018;

    b. **a guidance note on how to assess the adequacy of survey design** is produced, building on the experience of the MV team in verifying the Supplier's surveys; and

    c. **a guidance note on how to design a systems-based verification approach** is produced, so that future programmes can learn from the experiences of the MV team.

Ensuring sustainability

3. **Ensure strengthening local government is not marginalised over the coming two years:** Suppliers' efforts to engage with, support and develop the capacity of local governments and public utilities should intensify over the Outcome Phase. All Suppliers have to some extent worked with government agencies during the Output Phase, and such support and capacity development activities are a natural part of their work. An important strategic goal of the Outcome Phase should be to leave local governments significantly better able to support the services provided within their administrative under the WASH Results Programme jurisdictions than was previously the case. To achieve this, we recommend that Suppliers engage closely with local government to ensure a transfer of responsibility before programme end – for

---

[27] https://sustainabledevelopment.un.org/post2015/transformingourworld.

example, ensuring that local government carry out follow-up activities to sustain ODF achievements and to ensure the long-term functionality of water installations.

4. **Financing sustainability:** All Suppliers should undertake thorough analyses of the recurrent costs[28] of sustaining the water and sanitation services which they have implemented in the Output Phase. They should work with service users, and local and national governments to agree on the appropriate division of responsibility for cost-sharing. Such agreements should realistically take account of the ability of low-income service users to pay. Without a clear understanding of the likely magnitude of recurrent costs, and clear commitments regarding the responsibilities for their payment, it is likely that services will deteriorate over time.

5. **Clarify the advisory function of the MV team**, particularly with regard to advising Suppliers on the monitoring framework for sustainability. For SAWRP and SWIFT, where sustainability frameworks are not linked to verification, it is not clear to the Suppliers how recommendations by the MV team should be addressed.

9.2   For future WASH programmes (including with and without PBR)

This section is divided into two parts. First, a number of recommendations are made for large WASH programmes in general, with or without an element of PBR financing. Second, recommendations are made for why, when and how PBR financing should be used in such programmes.

9.2.1   Recommendations for large multi-country WASH programmes

Ensuring effectiveness

6. **Focus on equitable and inclusive outcomes:** The imperative of providing services for all and 'leaving no one behind' should be given priority, both for reasons of natural justice and in order to maximise health impacts. We recommend that future WASH programmes encourage approaches which deliver services to all members of targeted communities and administrative areas. Interventions will need to be carefully designed to address the root causes of exclusion. The equity focus should be clearly set out in the theory of change, for example by requiring achievement of ODF communities as an explicit intermediary step. Requiring monitoring data to be disaggregated by wealth quintiles and other dimensions of inequity is also key. DFID should be clear at programme start on which disaggregations are required, so that activities can be adequately designed and monitored.

7. **Make provision for addressing acute water supply needs:** We recommend that contracts allow for at least some support to water supply (as was done for SAWRP), so that if a programme finds itself occasionally working in communities where acute water supply problems constrain hygienic practices, these constraints can be resolved.

Ensuring sustainability

8. **Focus on sustainable service delivery:** It is important that the sustainability of services and behaviour changes is given priority in future WASH programmes. If this focus is not pursued, gains which are made, and the investments which allowed them to take place, risk being lost to 'slippage', attrition and decline. We specifically make the following recommendations:

---

[28] Recurrent costs can be disaggregated into WASH-cost categories of operating expenditure (OpEx), capital maintenance expenditure (CapManEx), and direct and indirect support.

a. Future WASH programmes should explicitly require that progress towards both outputs and outcomes should occur continuously, from the programme start. Having two phases labelled 'output' and 'outcome' should be avoided.

b. Explicit sustainability requirements should be made and action initiated from the start, for example supporting the formulation of district WASH plans and strengthening the capacity of local NGO and government partners.

c. Explicit measures to promote and enable sustainability should be included in the programme theory of change.

d. If possible, implementation periods of at least five to six years should be allowed for, as this increases the likelihood of results being sustainable.

9.2.2   Recommendations on the design of WASH PBR programmes

The recommendations below are primarily for donors, such as DFID, that are considering the use of PBR in future WASH programmes.

Ensuring relevance

9. **Clarify the purpose of using PBR:** While the potential benefits of PBR generally are well documented, it is important to set out in the programme design what the rationale is for using PBR in this case, and to define the added value that it should offer. The mechanism for how PBR is expected to deliver its intended benefits should be clearly set out in a theory of change. This will allow DFID to more easily check under which circumstances intended benefits have, or have, not been manifested, allowing important contributions to be made for sector learning on this topic.

Ensuring effectiveness

10. **Review the size of the PBR component:** Experience in this programme has shown that adoption of a 100% PBR modality results in significant risk transfer to suppliers, and can be associated with a high reporting burden, as additional milestones are added to ensure some cash flow in the early part of the programme. Experience with PBR in other sectors suggests that the value of PBR lies not in reimbursing the direct costs of implementation (as NGOs are already motivated to do this) but in linking PBR payments so as to incentivise additional dimensions, such as quality, sustainability or equity, which might otherwise be marginalised, albeit unintentionally. Given the low degree of control which suppliers have over the achievement of WASH behaviour change, we recommend a hybrid design where the bulk of supplier payments are grant-based, with a relatively small percentage used as an incentive for good performance.

11. **Be more prescriptive on which 'results' payments will be linked to:** The experience of this programme shows that if suppliers alone define which 'results' are linked to payment, there is a risk that efforts in support of sustainability and equity will be marginalised. Payments should be linked to the quantity as well as the quality of outputs, the inclusiveness of outputs (equity), to sustainability prerequisites and to the level of sustainability of usage achieved at programme end. The challenge lies in proposing indicators which are easily measurable and adequately cover a minimum of equity and sustainability prerequisites. Examples are given here and in the box below:

a. As is the case under the WASH Results Programme, work in under-served geographical areas should continue to be rewarded by allowing a higher unit price-per-beneficiary.

b. Bonus payments should be taken into consideration; these could reward survey results which confirm that equity targets and water-point functionality targets have been met after a certain period.

c. DFID should ensure capacity building and systems strengthening are not marginalised, by including them in sustainability requirements linked to payment criteria.

d. An explicit requirement for achieving ODF status should be integral to the sanitation component where feasible, with independent verification based on government criteria and procedures.

e. Some limits should be applied to supplier flexibility to transfer programme activities to areas where it is easier and/or cheaper to deliver the results, in order to ensure sufficient attention is given to equity considerations.

Box 13: Lessons from the health sector on how to incentivise the pursuit of equity using PBR

In the health sector, a 'pay for performance' approach has become a relatively common way of incentivising better provision of health care. Payments can be directed towards the users of the health service (rewarding use of a particular service) or towards the providers of the service (rewarding the provision of a particular service, or the quality of service provision). There are three key recommendations on how to incentivise equity:

- *Individuals can be targeted,* for example using eligibility criteria in conditional cash transfers. This is not directly relevant to WASH programmes, where community-wide access is sought. However, reaching poorer individuals can be rewarded by giving a bonus payment, to ensure that they are not excluded from such initiatives.

- *Poorer geographical areas can be explicitly targeted,* as was done by many of the projects under the WASH Results Programme. While less precise, this can ensure a relatively high coverage of poor and disadvantaged groups. Targeting such locations can be rewarded by allowing a higher unit price-per-beneficiary.

- *Pro-poor channels can be promoted.* For example, door-to-door hygiene promotion campaigns are more likely to reach poor and disadvantages groups than mass gatherings.

Once such approaches are in place, it is essential to monitor *whether* poor and disadvantages groups have in fact been reached. This requires disaggregating beneficiary data by gender, wealth groups and other dimensions, such as disability.

Adapted from: Gwatkin 2009.

---

12. **Introduce upside incentives:** The current funding modality punishes Suppliers if targets are not met (down-side incentive), but does not counterbalance this with bonuses for over-delivery (upside incentive). Experience from the health sector, however, confirms that PBR can be designed to create upside incentives; the box below explains under which circumstances upside or down-side incentives are most effective at improving service quality.

Box 14: Lessons from the health sector on how to incentivise 'quality' performance using PBR

In the health sector, incentivising performance has been increasingly linked to rewarding a certain quality of service, rather than rewarding only the provision of a service. Quantifiable quality checklists are needed which measure and reward specific components of quality. Such checklists are context-specific and can relate to the process of implementation as well as the quality of the service delivered.

In the health sector two different approaches have been used to incentivise quality performance through PBR:

- the **carrot-and-carrot method** consists of purchasing services and adding a bonus (for example, up to 25%) for the quality performance;

- the **carrot-and-stick method** entails purchasing services but detracting money in the case of bad quality performance. When using this method, one can inflate the carrots a bit, thereby ensuring a certain effect on the quality factor.

Experience from the health sector offers the following lessons:

- The carrot-and-carrot method is more advisable where the implementing agencies have **limited access to alternative sources of funds and may face cash-flow issues**.

- **Baseline levels of quality make a difference to the effectiveness of either approach**: when implementation quality is already fairly high at baseline, the carrot-and-stick method is more effective at rewarding the best-performing implementing agencies or facilities. When quality is low at baseline, the carrot-and-carrot method better protects basic implementing agencies' income, while penalising low-quality implementing agencies.

- The health sector has developed a whole series of quantifiable quality checklists and scorecards for assessing service provision. Essential lessons are that the quality criteria need to be realistic (monitoring data can be collected) and objectively verifiable. Quality checklists should also be updated regularly to incorporate lessons learned and to allow quality standards to be set progressively higher.

Adapted from: Fritsche *et al.* 2014.

13. **Include a substantial inception phase:** At the March 2016 Learning Event there was consensus among Suppliers, Verifiers and DFID regarding the benefits of a longer inception period: ensuring that verification requirements are clear before any implementation activities begin was seen as essential for clarifying verification and monitoring requirements and for allowing Suppliers to sufficiently resource their support to the verification process. However, SNV cautioned that sufficient flexibility needed to remain to allow verification requirements to be adjusted if necessary, to avoid an overly rigid verification system. This phase should be grant-funded, with no targets for output delivery.

14. **Appoint the verification provider before implementation begins and be more prescriptive on standards for verification requirements:** This type of PBR programme was new for both DFID and the Suppliers, and inevitably involved a lot of learning-by-doing for all, especially on the design of the monitoring, reporting and verification frameworks. The experience under this programme has shown that better defining verification requirements up front will enable suppliers and verifiers to develop better methodologies and operational plans, deploy the right type and number of personnel, and get up to speed faster once the programme begins. Guidelines on the evidence requirements for verifying WASH outputs, WASH outcomes and sustainability prerequisites should be drawn up before implementation begins. Appointing the verification provider before contracting suppliers is highly advised. The inception period will then be used to tailor evidence requirements to specific intervention types. It may be beneficial to separate the learning role from the verification role, to narrow the focus of the MV team.

15. **Explore options for reducing the verification burden on suppliers:** The reporting and verification requirements under this programme have been heavy, and are unlikely to be replicable in another programme; the human resource requirements alone would probably

preclude rolling out the current verification approach to multiple programmes. Experience under this programme suggests that the reporting burden can be reduced:

a. by using mobile-based monitoring for output and outcome results, though this may be less well-suited to process-based milestones; and

b. by taking the systems-based approach to verification further and – if verifiers are satisfied that a supplier's internal QA and monitoring systems are of a good standard – verifying only a portion of results for every milestone, though verifiers could reserve the right to make spot checks or demand access to data at any time. The Harmonised Approach to Cash Transfers (HACT)[29] system used by UN agencies for the funding partner organisations is a useful point of reference here.

---

[29] See https://undg.org/home/undg-mechanisms/business-operations-working-group/hact-advisory-committee/.

References

AR 2014. First Annual Review of the WASH Results Programme.

AR 2015. Second Annual Review of the WASH Results Programme.

Battye, F. (2015). 'Payment by Results in the UK: Progress to date and future directions for evaluation'. *Evaluation* 21(2):189–203.

Boyes-Watson, T. (2015). 'How the increasing use by DFID of Contracts and Payment by Results may affect INGOs.' Research Paper. Mango: London. www.mango.org.uk/getconnected/mangocampaigns/paymentbyresults.

Bertone, M.P. and Meesen, B. (2012). 'Studying the link between institutions and health system performance: A framework and an illustration with the analysis of two performance-based financing schemes in Burundi'. *Health Policy Plan*.

Boyes-Watson, T. (2015). 'How the increasing use by DFID of Contracts and Payment by Results may affect INGOs.' Research Paper. Mango: London. www.mango.org.uk/getconnected/mangocampaigns/paymentbyresults

Birdsall, N. and Savedoff, W. D. (2010) 'Cash on Delivery: A New Approach to Foreign Aid'. Publication of the Center For Global Development (CGD), Washington, D.C. URL: www.cgdev.org/publication/9781933286600-cash-delivery-new-approach-foreign-aid

Carter, R. (2016; in review). 'Can and should sanitation and hygiene programmes be expected to achieve health impacts?' (draft)

Carvalho, S. and White H. (1997). 'Combining the quantitative and qualitative approaches to poverty measurement and analysis', Technical Paper 366. The World Bank: Washington D.C.

Clist, P. and Dercon, S. (2014). 12 'Principles for Payment By Results (PBR) in International Development'. URL: www.gov.uk/dfid-research-outputs/12-principles-for-payment-by-results-pbr-in-international-development.

Clist, P. and Verschoor, P. (2014). 'The Conceptual Basis of Payment by Results'. University of East Anglia, Norwich, UK. URL: http://r4d.dfid.gov.uk/Output/195131/

DFID (2012). Water, Sanitation and Hygiene Portfolio Review March 2012. www.gov.uk/government/uploads/system/uploads/attachment_data/file/214187/DFID_20WASH_20Portfolio_20Review.pdf.

DFID (2013a). Water, Sanitation and Hygiene Evidence Paper, May 2013. URL: www.gov.uk/government/uploads/system/uploads/attachment_data/file/193656/WASH-evidence-paper-april2013.pdf.

DFID (2013b). Business Case and Intervention Summary, Water, Sanitation and Hygiene (WASH) Results Programme. URL: http://iati.dfid.gov.uk/iati_documents/4116096.odt.

DFID (2014a). Sharpening incentives to perform: DFID's Strategy for Payment by Results.

DFID (2014b). Designing and Delivering Payment by Results Programmes: Short Technical Guidance. July 2014. URL: www.gov.uk/government/uploads/system/uploads/attachment_data/file/352519/Designing-Delivering-PbR-Programmes.pdf.

DFID (2014c). First Annual Review of the WASH Results Programme. URL: http://iati.dfid.gov.uk/iati_documents/4571521.odt.

DFID (2015a). Designing and Delivering Payment by Results Programmes: Full Technical Guidance (unpublished document).

DFID (2015b). Second Annual Review of the WASH Results Programme. URL: http://iati.dfid.gov.uk/iati_documents/4979606.odt.

DFID (2015c). DFID's Evaluation Framework for Payment by Results. URL: www.gov.uk/government/publications/dfids-evaluation-framework-for-payment-by-results.

DFID (2016). Third Annual Review of the WASH Results Programme. URL: http://iati.dfid.gov.uk/iati_documents/5498968.odt.

Engel, J and S. Nicolai (2012). 'Using case studies to untangle complexity and learn from progress'. ODI Development Progress. Project Note 2. www.developmentprogress.org/using-case-studies-untangle-complexity-and-learn-progress-0'

Fritsche, G., Bèla, R., Soeters, and Meessen, B. (2014). Performance-Based Financing Toolkit. Washington, DC: World Bank.

Gwatkin, D. R. (2009). 'Ensuring that the Poor Share Fully in the Benefits of Results-Based Financing Programs in Health'. World Bank Working Paper

Hunter, P.R. and Prüss-Ustün, A. (2016). 'Have We Substantially Underestimated the Impact of Improved Sanitation Coverage on Child Health? A Generalized Additive Model Panel Analysis of Global Data on Child Mortality and Malnutrition'. *PLoS ONE* 11(10): e0164571.

Lindenenauer *et al.* (2007). 'Public reporting and pay for performance in hospital quality improvement'. *New England Journal of Medicine* 356(5): 486–496.

NAO (2015). 'Payment by results: Analytical framework for decision-makers. A review of National Audit Office'. London, UK.

Perrin, B. (2013). 'Evaluation of Payment by Results (PBR): Current Approaches, Future Needs'. Working Paper 39. DFID, London.

Quinn Patton, M. (2015). *Qualitative research and evaluation methods: Integrating theory and practice.* Sage, Fourth edition.

Sida (2015). 'Results Based Financing Approaches (RBFA) – What are they? A methodological introduction'. URL: www.sida.se/English/publications/Publication_database/publications-by-year1/2015/november/results-based-financing-approaches-rbfa--what-are-they/

Stern *et al.* (2012) Broadening the Range of Designs and Methods for Impact Evaluations. Working Paper 38. DFID. London, UK. www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

White, H. and Phillips D. (2012). 'Addressing attribution of cause and effect in small impact evaluation: towards an integrated framework,' 3ie Working Paper 15.

Annex A   Terms of reference and qualifications to the terms of reference

This annex provides the terms of reference for the MVE component, and summarises the clarifications received from DFID on the terms of reference relating to the Evaluation component.

A.1   Original terms of reference

Introduction

1. DFID is seeking a service provider to provide an independent and rigorous monitoring, verification and evaluation (MVE) function for the WASH Results Programme (a large delivery programme providing access to water, sanitation and hygiene).

2. These Terms of Reference (TORs) set out DFID's requirements for monitoring, verification and evaluation including the objective, scope of works and outputs for the function and other relevant design information.

Objective

3. To provide an independent and rigorous monitoring, verification and evaluation function for the WASH Results Programme.

Recipient

4. The recipient of this service will be the DFID WASH Policy Team.

Background

5. The Department for International Development (DFID) manages the UK's aid to poor countries and works to get rid of extreme poverty. DFID is working to reach the Millennium Development Goals (MDGs), the international targets agreed by the United Nations (UN) to halve world poverty by 2015.

6. Progress on Water, Sanitation and Hygiene (WASH) provision is critical to the achievement of these targets. Millennium Development Goal (MDG) 7c specifically relates to sustainable access to safe drinking water and basic sanitation by 2015.

7. The UK Government has committed to scale up our results on Water, Sanitation and Hygiene to support 60 million people gain access to sustainable WASH services before the end of December 2015. To achieve these results, DFID has commissioned a portfolio of programmes, including a WASH Results Programme managed by the WASH Policy Team in Human Development Department of DFID's Policy Division.

8. The WASH Results Programme is a results oriented fund which provides funding to organisations capable of delivering water, sanitation and hygiene interventions for people who currently do not have access to clean water or safe sanitation. WASH Results Programme projects must demonstrate real, positive, sustainable changes to the lives of

poor people. DFID will award an expected three contracts[30] to organisations based on payments linked to output results (payment by results).

9. WASH interventions supported by DFID must be aimed at providing poor people with access to improved drinking water supplies, access and use of basic sanitation and the adoption of behaviours that reduce the health risks caused by poor hygiene, including hand washing at critical times. This support should enable approximately 4.5 million people to receive access to sustainable WASH services by December 2015. The WASH Results Programme will be managed within DFID with contracts reporting to an identified fund manager within DFID. Full details of the WASH Results Programme will be available on DFID's website www.gov.uk/dfid

10. DFID' approach[31] to payment by results involves three key elements:

    i.   Disbursements tied to the achievement of clearly specified results: payment for outcomes such as completion of education, rather than payment for inputs such as provision of textbooks;

    ii. Recipient discretion – the recipient has space to decide how results are achieved; and

    iii. Robust verification of results as the trigger for disbursement.

    These relate directly to this results programme: i. the TORs for the overall programme set the results expected; ii, organisational discretion with the successful contract holder on the approach to reaching the required results, and iii. robust verification, which for this programme includes strong systems in the contracted organisations for achieving the results, and the services covered in these terms of reference to independently audit the quality of the contracted organisations results reporting.

11. The monitoring and verification of the results achieved through the WASH Results programme will be critical to the credibility and robustness of the data used to track and report on progress. To support the DFID Fund Manager in ensuring robust monitoring, verification and evaluation systems and data are in place, we are now looking to appoint a Monitoring, Verification and Evaluation (MVE) Service Provider to run for the lifetime of the WASH Results Programme over the next 5 years.

Scope of Work and Requirements

12. Design and implement a system of third-party results verification (including working with WASH Results Programme Suppliers to define measurement methods and monitoring, verification and reporting requirements for payment) for those parts of projects operating under a Payment by Results framework. This will be used to verify results and outputs to inform payments. The MVE provider will audit contractors' results reporting within the contract terms related to payment by results.

13. The MVE Service Provider will develop a monitoring framework to assess progress and performance of individual projects funded under the WASH Results programme. Where this indicates that corrective action is required, the MVE Service Provider will inform both DFID and the implementing partner and facilitate discussion of corrective actions required. The MVE Services Provider will also assess and audit the monitoring systems and results reporting within each project and make recommendations for improvement where required. Assist DFID with monitoring contracts under this Results Programme. These monitoring tasks will include:

    -   early monitoring of activities undertaken in the inception phase of the main results

---

[30] As per the TORs for the main programme, 3 contracts are expected (two WASH contracts each to reach 1 million people, and one sanitation and hygiene contract to reach 2.5 million people). However, DFID reserves the right to award more contracts (up to a maximum of five) or fewer contracts, as outlined in the TORs.

[31] https://www.gov.uk/government/publications/payment-by-results

contracts

- Un-announced visits to project work.

- Quality assurance of contract holders results returns

- Facilitation of annual review missions with DFID involvement.

14. Develop one or more impact evaluations on key issues agreed during the inception with DFID, but potentially covering one or more of the following: approaches to service delivery at scale, efficiency and effectiveness across different contexts, sustainability and Payment by Results.

15. The MVE Service Provider has knowledge and understanding of:

- monitoring and evaluation of development programmes using both quantitative and qualitative methods;

- monitoring and evaluation of WASH programmes, including testing of WASH outcomes;

- social research management;

- Payment by Results programmes

- WASH sector governance and political economy analysis.

- management of impact evaluations; and

- evaluations in the context of major donor interventions, ideally focused outside of government.

16. In addition the MVE Service Provider is expected to:

- establish a good working relationship with the DFID Fund Manager and implementing partners;

- support the DFID Fund Manager to establish appropriate monthly reporting mechanisms;

- provide technical support on the use of Payment by Results;

17. The services under the monitoring and verification are primarily to support DFID's role in holding results contract holders to account for delivery within their contracts. Timeliness of work under the monitoring and verification is therefore critical. The evaluation component will also provide findings to inform annual reviews of the programme, to allow for learning-feedback to inform progress under the project and permit course correction.

18. Where possible the evaluation questions, identified in the inception phase, should follow the OECD DAC Criteria for Evaluating Development Assistance[32] including the headings of relevance, effectiveness, efficiency, impact and sustainability.

Outputs

19. The MVE Service Provider is responsible for drafting a Monitoring, Verification and Evaluation Framework for approval by DFID within the first 3 months. This inception report should contain:

- a monitoring, verification and evaluation strategy and implementation plan;

- a risk management plan;

- a quality assurance plan;

- an outline of proposed methods for assessing core indicators;

---

[32] http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

- an outline of the proposed approach to assessing project-specific additional indicators;

- an outline of the proposed approach to assessing programme performance;

- an outline of proposed evaluation questions (based on identification of what is feasible within given constraints), design and methods; and

- a strategy for disseminating information to key stakeholder and partners.

20. Once the inception report it is approved, it is expected that the MVE Service Provider will be responsible for delivering the outputs outlined below.

21. Support technical review of contractors proposed work, including assessment of the approaches proposed, review of contractors approach to sustainability, and checks on environment appraisals.

22. Tracking progress to ensure robust measurements of performance at the project and programme level including:

- auditing partner approaches to monitoring and carrying out reviews, field-work and spot checks to provide assurance of robust performance management processes;

- establishing data requirements to inform value for money and sustainability metrics. Value for money metrics will include key unit costs (eg. cost per latrine, cost per person of providing improved water supply) and also measures of effectiveness (measures of behaviour change such as the percentage of people reported as benefitting from hygiene interventions that are hand washing with soap and at critical times or percentage of beneficiaries that continue to use improve sanitation facilities). DFID has provided further information[33] on understanding value for money at the economy, efficiency and effectiveness levels. Sustainability metrics will include system functionality, coverage of lifetime costs and other key factors likely to promote sustainability such as institutional capacity and environmental factors. Programme monitoring data and additional spot checks will inform these metrics.

- allowing comparison between programmes and to inform future DFID and wider sector programming;

- quality assuring project progress reports, with a focus on ensuring robust tracking of results and performance based on agreed milestones and targets and challenging data and conclusions if necessary;

- notifying DFID of progress with projects at least quarterly and monitoring, managing and advising DFID on risks to programme delivery and mitigation measures;

- aggregating data and preparing programme progress reports to provide overall assessments of programme performance against milestones and targets in the WASH Results Programme logframe; and

- aggregating end of project data and preparing a programme completion report.

23. The Monitoring, Verification and Evaluation Service Provider will also be responsible for verifying outputs, by auditing the reports of service providers, to inform results-based payments. This will include:

- Working with contract holders to establish results baselines

- Undertaking assessments of organisations capacity to report on results

- Advising organisations' on improvements to process for results reporting.

- Provide comments (by desk review) on the quarterly reports from service providers.

---

[33] www.gov.uk/government/publications/dfids-approach-to-value-for-money-vfm

- Provide a full audit, annually, of contract holders' results reporting, including visiting projects.

24. In early 2016 the DFID Fund Manager will arrange an event at which the WASH Results Programme projects will be able to demonstrate the main results, findings and lessons from programme funded activities. The MVE services provider will support this by preparing a document and a presentation which summarises the results to-date and the lessons learned.

Evaluation Requirements

25. Development of a detailed strategy and implementation plan for evaluation including evaluation design and methods and timings for collection of baseline data.

26. Work with DFID to select, design and administer one or more depth evaluations at programme level and on a select number of project interventions and thematic areas. These decisions will be based on relevance to the overall objectives of the WASH Results Programme, potential for wider DFID and global lesson learning and the potential to fill key knowledge gaps and feasibility and cost of collecting data.

27. Implementation of the evaluation framework agreed on with DFID including:

- tracking whether assumptions set out in the Theory of Change and logframe hold;

- evaluating innovative elements of projects (including results-based finance);

- developing the evidence base on value for money metrics, comparing Suppliers and approaches across contexts;

- identifying factors which have enhanced or impeded the sustainability of WASH interventions;

- undertaking case studies of WASH projects at scale, to support the evaluations.

- addressing knowledge gaps (e.g. effective approaches to behaviour change, community monitoring and payment by results); and

- assessing how the wider environment has enabled or impeded achievement of programme objectives and identifying implications for programming.

28. Dissemination of lessons to inform WASH Results Programme evolution and wider DFID and global programming including:

- with the DFID Fund manager, disseminating lessons learned and reporting those to DFID to agree actions for the WASH Results Programme;

- ensure lessons are drawn out on what works in WASH, triangulated with other evidence, and reported to DFID. These lessons may be both immediate and used to inform future WASH Results Programme evolution or longer term and used to inform future DFID or others' interventions;

- ensure lessons are drawn out on the costs and benefits of the Results Programme approach compared to other approaches including DFID bilateral aid and other DFID operations;

- linking with the WASH Operational Research Programme to share monitoring information and evaluation findings and to co-ordinate analysis and dissemination;

- design an innovative strategy to disseminate lessons through a variety of mediums and engage key partners and stakeholder in lesson learning on implementation and good practice from the WASH Results Programme, ensuring this reaches a wide audience (including project implementing partners, national governments, DFID country offices, the private sector and civil society).

Constraints and Dependencies

29. The MVE Services Provider will be expected to provide its own overseas duty of care in relation to its employees and other personnel it retains and logistical arrangements. If deemed necessary DFID may need to be convinced that systems and procedures that it has in place are adequate if traveling to conflict affected countries.

30. Clear separation between the team implementing the monitoring and verification elements of these services and the team implementing evaluation will be required to ensure independence of the evaluation elements.

Reporting and Monitoring and Evaluation of the MVE Services Provider

31. Key Performance Indicators (KPIs) will be agreed between the MVE Services Provider and DFID during the Inception phase. These will ensure that the management of the contract is undertaken as transparently as possible and to ensure that there is clarity of roles and responsibilities between the DFID WASH Results Programme team and the MVE Services Provider. The MVE Services Provider will need to demonstrate to DFID, at quarterly intervals its performance against these KPI's.

32. DFID will evaluate the performance of the MVE Services Provider throughout the life of the programme and at least twice yearly including as part of DFID's standard Annual Review of the programme. The MVE Services Provider will be expected to submit progress reports and lessons presented written and orally to DFID twice annually in-line with DFID's programme cycle as outlined in the requirements section of this ToR. It is expected that the MVE Services Provider take a proactive approach to notifying DFID of any matters which may require immediate attention.

33. The Monitoring, Verification and Evaluation components will report directly to DFID. To ensure that the evaluation (which would include evaluation of DFID's role in fund management of the programme) is rigorous, all evaluations will be made public and subject to the scrutiny which DFID operates within, which includes parliament and the Independent Commission on Aid Impact.

Timeframe and Contractual Arrangements

34. The contract for the MVE Services Provider will be awarded from May 2014 to November 2018. The contract is designed to end after financing is dispersed to allow a final evaluation of projects to be completed. The inception phase will be for a period of 6 months. DFID reserves the right to extend the contract for a further 2 years, on basis of continued need, and availability of funding.

35. The WASH Results Programme will run for 4.5 years (2014 – 2018).

36. DFID will issue a contract for the full programme duration however there will be a formal break clause in the contract at the end of the inception period. Progression to the implementation phase will be dependent on strong performance by the SP during the inception period and delivery of all inception outputs, including a revised proposal for implementation period. We expect costs for implementation to remain in line with what was indicated in your original proposal (any increase in costs will be subject to agreement with DFID), with costs such as fee rates fixed for contract duration. DFID reserves the right to terminate the contract after the inception phase if it cannot reach agreement with the SP on the activities, staffing, budget and timelines for the implementation phase.

37. DFID reserves the right to scale back or discontinue this programme at any point (in line with our Terms and Conditions) if it is not achieving the results anticipated. Conversely, we may also scale up the programme should it prove to be having a strong impact and has the

potential to yield better results.

DFID coordination and management

38. The DFID WASH Fund Manager will have the day-to-day oversight and management of the MVE Services Provider. The DFID WASH Fund Manager will monitor operational and financial progress drawing on inputs from the WASH Team staff and will raise any issues that require attention to DFID senior management and Ministers as necessary.

39. The DFID Evaluation Department (EvD) support the DFID Fund Manager by providing strategic advice as required and ensuring that evaluation and monitoring activity undertaken by the MVE Services Provider aligns with wider DFID activity. The DFID Fund Manager team will work alongside the MVE Services Provider to consider what input is required, by whom and at what times to ensure technical advice is on hand at the right times.

40. There will be regular meetings between the DFID Fund Manager, staff from the WASH Policy team and the MVE Services Provider.

A.2   Clarifications to the terms of reference

This section clarifies our understanding of the terms of reference for this evaluation. The terms of reference specify developing '*one or more impact evaluations on key issues agreed during the inception with DFID, but potentially covering one or more of the following key issues agreed during the inception with DFID: approaches to service delivery at scale, efficiency and effectiveness across different contexts, sustainability and Payment by Results*' (see Annex A.1 for the full terms of reference).

We propose to address a selection of these issues through activities across two interlinked workstreams – namely the **PEW and the RCT Research Study.** Based on the list outlined in the terms of reference, these will jointly gather evidence on 'Payment by Results', 'efficiency and effectiveness across different contexts' and 'sustainability'.

With regard to the evaluation purpose

Item 17 of the terms of reference state that the evaluation component '*will […] provide findings to inform annual reviews of the programme, to allow for learning-feedback to inform progress under the project and permit course correction*'. Item 27 also highlights that the evaluation framework would '*address knowledge gaps (e.g. effective approaches to behaviour change, community monitoring and payment by results)*'. We therefore understand the evaluation to be primarily undertaken for learning purposes, as opposed to accountability purposes. As a result, our evaluation design favours depth over breadth: we propose a case study approach to understand the mechanisms operating under certain contexts, as opposed to an evaluation design which is able to verify programme outcomes in every programme country.

Nonetheless, for practical reasons, the dissemination of the evaluation findings (i) to the Suppliers, (ii) to DFID, and (iii) to a wider policy audience will be organised under the L&D component managed by Itad. This choice was made because the verification team is already planning regular feedback and learning sessions with DFID and the Suppliers. Instead of duplicating efforts, the evaluation team will participate at these meetings to disseminate the findings of the evaluation.

With regard to the evaluation scope

Item 26 of the terms of reference outlined that the evaluation team would select, design and administer one or more in-depth evaluations at programme level on a select number of project interventions and thematic areas. As a result, we propose both in-depth evaluation activities through country case studies, as well as broader evaluation activities at programme level.

With regard to the RCT Research Study

The Organisation for Economic Co-operation and Development Development Assistance Committee (OECD–DAC) criteria define 'impact' as the causal effect of the programme on indicators of interest. The programme's effect can relate to medium-term 'outcomes' (e.g. use of water or sanitation services) or to longer-term 'impacts' (e.g. improved health and welfare as a result of the use of water or sanitation services). In both cases we would refer to these as 'impact' evaluations.

While the delivery of programme outputs should generally lie within the control of Suppliers, outcomes and impacts lie progressively further from their control. The achievement of desirable **outcomes** involves the use of services and the practice of new behaviours. There is a level of intent and choice on the part of service users and those practising improved hygiene behaviours, which Suppliers cannot fully guarantee – although they can and should influence those choices.

Beneficial **impacts** on health may require concurrent changes in nutrition, in the cleanliness of the living environment, and in the behaviours of others (e.g. community-wide compliance with latrine use and good hygiene practice), amongst other factors. Similarly, impacts on educational attainment, income and quality of life only follow if numerous other conditions are fulfilled. These mostly lie outside the sphere of influence of WASH implementing agencies. Better WASH outputs and outcomes are necessary but not sufficient conditions for the achievement of beneficial impacts.

In short, there are four main reasons why our evaluation will not focus on attributing any changes in longer-term health or non-health 'impacts' to the WASH Results Programme:

1. impact indicators (especially those related to health and quality of life) are difficult and expensive to measure – undertaking such measurement would not represent VFM;

2. we currently possess inadequate theories of change and explanatory capability to say definitively why certain impacts have or have not materialised – we may find that diarrhoea morbidity in under-fives has reduced by a certain percentage, but we cannot explain why it was not reduced by more than this; or we find no change in some other indicator, but can only speculate as to why;

3. we already know that outcome-level results in WASH are some of the preconditions for beneficial impacts – we would learn nothing new; and

4. the use of better WASH services and the practice of better hygiene are worthwhile results in themselves, and they should be available to all.

As a result, in consultation with DFID, we have proposed the 'impact evaluation' – referred to as RCT Research Study in this document – to focus our evaluation on medium-term 'outcomes', which would focus on the following aspects set out in the terms of reference:

- '*identifying factors which have enhanced or impeded the <u>sustainability</u> of WASH interventions*'; and

- '*assessing how the <u>wider operating environment has enabled or impeded</u> achievement of programme objectives and identifying implications for programming*'.

The likelihood of health and non-health impacts being achieved by the WASH Results Programme will nonetheless be explored in a qualitative manner, to better understand the prerequisites which favour health impacts, and to understand if there have been any unintended positive or negative impacts on service users as a result of the WASH Results Programme.

With regard to the PEW

The programme evaluation will focus on two of the aspects listed in the terms of reference:

- '*Efficiency and effectiveness across different contexts: develop the evidence base on value for money metrics, comparing Suppliers and approaches across contexts*'; and
- '*Evaluate <u>innovative elements</u> of projects (i.e. the results-based finance modality).*'

We understand the former to be part of the tracking of VFM metrics, listed under *verification* activities in the terms of reference. However, the terms of reference assume that programme monitoring data and additional spot checks will be sufficient to inform these VFM metrics. As this is not the case, VFM analysis will primarily fall under the responsibility of the *evaluation* team, instead of the verification team. Due to the additional cost of complementing the insufficient monitoring data, this VFM analysis will be undertaken in case studies, as opposed to in every programme country.

With regard to the latter, we will examine the role of the PBR modality in delivering outcomes through a theory-based approach.

With regard to the evaluation questions

The terms of reference did not outline any evaluation questions, proposing that these, where possible '*should follow the OECD DAC Criteria for Evaluating Development Assistance including the headings of relevance, effectiveness, efficiency, impact and sustainability*'. As a result, we have drafted our own HEQs in line with the DAC criteria, The HEQs were signed off by DFID in September 2014.

Annex B   Theory of change of the programme

This annex presents an explicit articulation of the WASH Results Programme's theory of change and its underlying assumptions. These form the basis for this evaluation, with the evaluation questions set out in Chapter 3 testing key assumptions underlying the theory of change.

For evaluation purposes, we distinguish theories of change at two levels. Firstly, a programme-wide theory of change articulates how the WASH Results Programme contributes to a chain of overarching, WASH-related outputs and outcomes that seek to have an impact on the lives of beneficiaries in the WASH Results Programme targeted communities. This theory of change is presented in Section B.1. Secondly, nested within this programme-wide theory of change we identify a theory of change of the PBR modality, which articulates how the PBR modality contributes to changes in the behaviour of the service providers directly influenced by the PBR modality. This second theory of change is presented in Section B.2. We consider the latter to be nested in the former because the PBR modality, in combination with other programme features, is expected to ultimately influence outputs and outcomes at beneficiary level because of the changed behaviour of the service providers.

The distinction made betweentheories of change at two levels is purposeful, in the sense that it is a heuristic tool to support the objectives of the evaluation. On the one hand, the evaluation is meant to assess the influence of the PBR modality, which requires us to assess change and mechanisms of change from the perspective of the PBR modality. On the other hand, the evaluation needs to answer whether the programme successfully achieved its WASH-related outcomes, which requires a programme-wide perspective.

## B.1   The WASH Results Programme's theory of change

Figure 12 presents a modified version of the WASH Results Programme's theory of change. This was developed following a critical assessment of the original theory of change, a detailed review of the project design documentation, and discussion with DFID. We have included elements that were not explicitly articulated in the original theory of change, but that were set out in other DFID documentation (in particular the Business Case). The revised version of the theory of change was discussed together with DFID in February 2015 and revised accordingly.

Table 16 below sets out all the explicit assumptions that underpin the theory of change. It is not an exhaustive list of all the assumptions which are *implied* by the theory of change (but not stated in any documentation). It includes assumptions stated in the Business Case (the item number is indicated), as well as DFID's evidence papers. The assumptions relating to PBR are not included in the table, as these are discussed separately in Section B.2.

The theory of change forms the basis for this evaluation, with the evaluation questions developed to test the key assumptions that underpin it (explicit and implicit). It is important to note that the programme-wide theory of change remains generic in nature, without a detailed explanation of how specific service providers' interventions will achieve the programme outcomes. This is a logical consequence of the PBR modality of the programme, which allows the service provider flexibility in regard to how best to achieve outputs and outcomes given the implementation context.

Figure 12:   The WASH Results Programme theory of change

Boxes shaded in red represent items which were not articulated in the original theory of change. Red text in italics represents wording used in DFID documentation but also not articulated in the original theory of change.
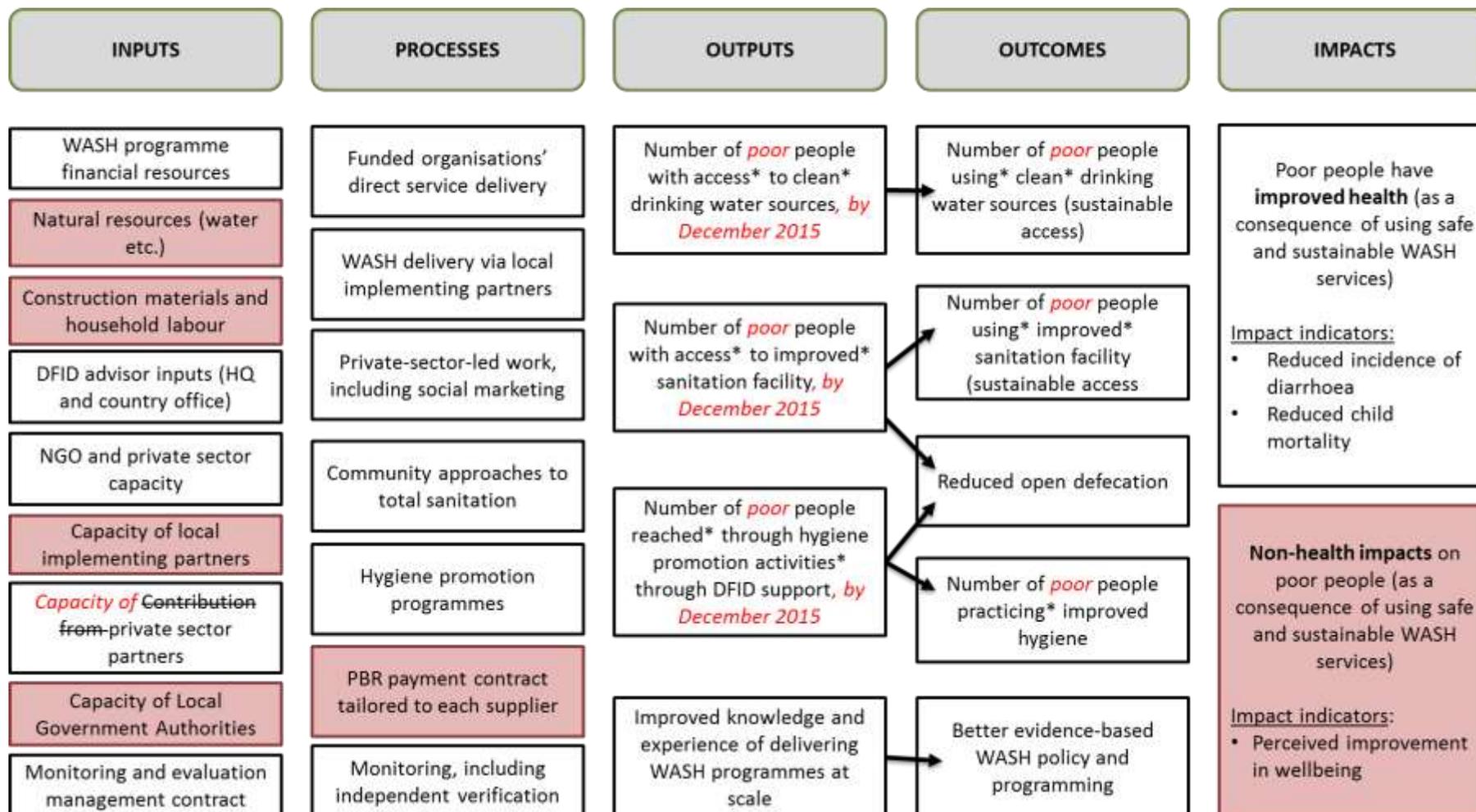
(*) These items were defined differently by each Supplier consortia, as articulated in their M&E framework.

**Table 16:**   **Assumptions stated to be underlying DFID's theory of change**

Inputs to processes	Processes to outputs	Outputs to outcomes	Outcomes to impact
• Sufficient engagement, level of expertise, continuity and speed of response by DFID adviser (#53) • MVE contract enables 'prompt corrective action' (#124) • Sufficient user demand for services (and capacity of the community to manage improved services) (evidence papers)	• Any construction and capacity development activities are appropriate and of high quality • Vulnerable groups are included in programme (#23) • Households make labour and capital available (#37) • Private sector has capacity to construct and rehabilitate water-points (#37) • Communities are able to operate and maintain the improved water supply (#37) • Local government authorities have the capacity to manage inputs and deliver outputs (#37) • Hydrogeological, hydrological and water resources management assessments have been undertaken (#68) • Climate risk assessments undertaken (#78, 79, 80) • Sustainability assessments undertaken (#82) • Effective targeting of poor/vulnerable groups (#117–121)	• Households change their behaviours as a result of sanitation and hygiene campaigns (#37) • Life-cycle costs and responsibilities for their payment have been included (#15) • Community participation and capacity (#15, #16) • No impact of time-bound PBR targets on sustainability (#117–121)	• A critical mass of households change their behaviours, resulting in health and non-health benefits at community level (evidence papers) • No unsustainable abstraction of water or pollution from sanitation (#68) • No use of contaminated water sources (#68) • No contamination of water after collection from a 'safe' source (evidence papers)

B.2 The PBR modality theory of change

An important feature of the WASH Results Programme is the PBR modality embedded in the programme. This modality is meant to influence service provider behaviour, and therefore is based on its own hypothesis about how service provider behaviour is assumed to change. The original WASH Results Programme theory of change did not make explicit reference to a PBR modality. We have included the PBR modality as an element in the modified WASH Results Programme theory of change (see Figure 12, above), complemented by an explicit nested theory of change for the PBR modality (see Figure 13). We use DFID's generic guidance on designing PBR programmes (DFID 2014) as a basis for understanding the rationale for using a PBR modality for the WASH Results Programme, and the associated assumptions underpinning it.

The figure below presents DFID's generic theory of change for PBR programmes. In the context of the WASH Results Programme, the recipient is the service provider, i.e. the lead organisation for each of the three Supplier consortia.

Figure 13:   DFID's generic theory of change for a PBR programme
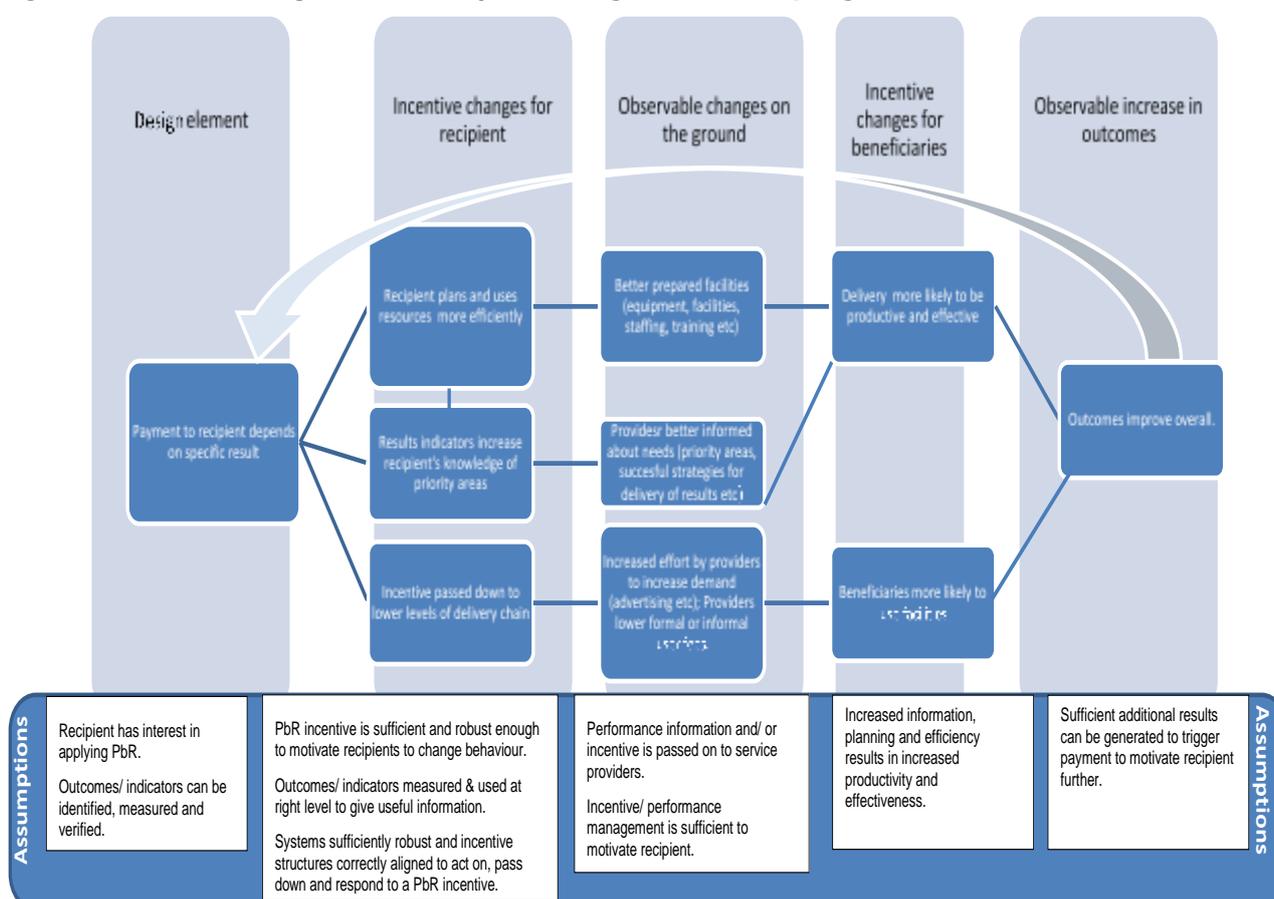
In addition to the critical assumptions highlighted in the PBR theory of change diagram above, DFID's guidance notes highlight further assumptions and necessary preconditions for successful implementation. In the context of the WASH Results Programme, these assumptions translate as follows:

B.2.1 Design assumptions

With regard to inputs and processes

- Service providers' incentives are perceived as the main constraint to achieving the scale of MDG-related WASH outcomes that the WASH Results Programme aims to achieve.
  - ο PBR will be an effective tool for shifting the incentive frameworks within service providers (predominantly INGOs).
  - ο Inefficiencies in INGO practices are perceived as a key constraint to reducing the unit cost of the MDG-related WASH outcomes that the WASH Results Programme aims to achieve.
  - ο PBR is an effective tool for increasing the efficiency of INGOs in delivering WASH outcomes.
- PBR (which imposes a risk on Suppliers) is consistent with DFID's stated desire for innovation in the delivery of WASH services (assuming innovation entails more risk).
- The organisations expected to bid for the implementation of the WASH Results Programme have previous experience of operating under a PBR contractual framework and fully understand the operational and financial implications.
- The Suppliers (INGOs) have the ability to pre-finance large-scale programmes.
- The contracting agent (DFID) has sufficient capacity for designing, procuring and managing large-scale PBR projects.
- The PBR design does not negatively affect the choice of countries covered by the programme (i.e. Suppliers avoid working in risky countries).
- The PBR modality will not discourage effective engagement with national government institutions and systems, at local, regional and national levels.

With regard to outputs, outcomes and impact

- The programme's intended outputs and outcomes can be accurately identified, measured and verified (items #117–121 in the Business Case).
- Suppliers (INGOs) have direct control over the intended outputs and outcomes (both in terms of external factors, but also in terms of traditional sub-contracted delivery systems).
- The additional benefits (i.e. achievement of otherwise unattainable outcomes and/or lower unit costs per outcome, additional transparency and accountability, and potential for delivering higher quality, equitable and sustainable outcomes) are expected to outweigh the costs (both direct and indirect) of the PBR modality (i.e. designing effective PBR contracts and incentive frameworks, payments frameworks schedule and verification systems).
- The PBR framework can be designed such that it does not incentivise achievement of payment-triggering results at the expense of long-term sustainability in WASH outcomes.

Implementation assumptions

- The 'results' against which the implementing partners' performance is assessed are designed effectively to correspond to the WASH Results Programme's intended outcomes (plus activity and outputs where relevant).
- The links between achievement of results and payments are clearly defined and appropriate (e.g. indicators, payment levels).
- Where the PBR framework involves explicit performance incentives (i.e. a pro-rata payments system, bonus payments for over-achievement and/or penalisation for under-achievement), the performance indicators and payment level chosen are appropriate.

- The 'results' to be measured and the payments triggering process are clearly defined in the implementing partners' contracts with DFID.
- There is a clearly defined process for contesting payment trigger decisions.
- The verification reporting processes have been designed effectively (and in a timely manner) so as to ensure appropriate and efficient reporting demands on implementing partners, and based on existing information wherever possible:
  - ο the verification cycle timeframe is consistent with contractual obligation and Suppliers' cash-flow constraints; and
  - ο the verification process is accurate and sensitive to local community contexts.
- Suppliers engage with the verification process (in an observational role or otherwise) to encourage accuracy, buy-in and usefulness of verification information, and to facilitate access by the verification teams.
- The financial risk associated with partial or non-payment for under-achievement is shared (fully or partially) between the consortium lead organisation and national implementing partners within each consortium.
- The PBR modality does not result in perverse incentives, resulting in the design and implementation of sub-optimal programme interventions (e.g. less risky).
- The PBR modality does not encourage gaming (e.g. deliberately improving performance measures).

Annex C   Detail on evaluation approach

C.1   Evaluation matrix

DEQs	Our current conclusion	Data sources	Methodology / basis for evaluation judgement	Report section	Confidence level
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?	Type and scale of anticipated results were clearly defined for outputs, but level of ambition at outcome level was much less clear at the design stage. Objectives relating to equity and sustainability were stated but their relationship to the PBR modality was not explicit in the Output Phase.	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation • WASH best practice review	Evidence of clear articulation of all of DFID's underlying programme objectives in the Business Case, the ITT and other DFID documentation. This was triangulated with the views of DFID, Suppliers, and unsuccessful bidders. Consistency of stated objectives between documents.	4.1, 4.2, 4.3	HIGH
DEQ 1.2: To what extent does the **programme's design** (i.e. the theory of change) set out a clear and realistic process for how programme activities will achieve the intended outputs, outcomes and impacts?	The theory of change lacked important detail: assumptions were not clearly set out, and the rationale for, and contribution of, PBR was not explained.	**Primary** • DFID interviews • Lead Supplier interviews **Secondary** • DFID documentation	Evidence of consistency between DFID interviews and DFID documentation, triangulated with the views of DFID and Suppliers. Evidence of completeness of theory of change assumptions in DFID documentation.	4.4	HIGH
DEQ 1.3: To what extent were the **scale and pace of the programme** (including the Dec. 2015 deadline) realistic for achieving intended outputs and outcomes given the capacity	Suppliers explained how/why they set realistic targets and selected partners. However, the urgency of the Dec. 2015 deadline in some cases affected quality and sustainability of results.	**Primary** • DFID interviews • Lead Supplier interviews	Evidence in the ITT of tailoring of WASH Results Programme design to capacity and context, triangulated with the views of DFID, Suppliers, and unsuccessful bidders.	4.6	HIGH

of Suppliers and their local partners?		**Primary** • Unsuccessful bidder interviews **Secondary** • DFID documentation			
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?	Rationale for using PBR unclear at start, but DFID has clarified that priorities were to stimulate the formation of new partnerships for delivery at scale and to achieve greater certainty around the delivery of results (quantity and timing). These were achieved, but PBR made Suppliers risk-averse in setting targets. Incentives were framed negatively in original design: penalties for not meeting targets but no bonuses for over-delivery. PBR modality not wholly appropriate as equity / sustainability aspects were not built into payment criteria. No official inception phase for Suppliers but considerable inception work needed to design MV systems given that programme design was new for DFID and Suppliers. Government engagement is important for sustainability but it was challenging to meet output deadlines via government agencies that are not subject to PBR.	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews • LV interviews **Secondary** • DFID documentation • Verification Design Document	Evidence of the expected causal link between PBR and achieving the objectives in programme documentation and verification methodology, triangulated with the views of DFID, Suppliers, Verifiers and unsuccessful bidders. Perceptions of sufficient capacity of DFID and expected bidders to implement a PBR programme (triangulated across different interviewees).	4.7 5.1	**HIGH** Minor risk of obtaining different opinions within DFID
DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?	Some private sector bidders were put off by uncertainty around how the PBR modality would operate.	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation • WASH best practice review	Evidence of clear intent of the programme to encourage 'innovative' private sector partnerships, based on DFID documentation. Assessment of likelihood of private sector partnerships based on triangulation of perceptions of DFID, Suppliers, and unsuccessful bidders.	4.8.1	**HIGH**

DEQ 1.6: How likely was it that the programme design would encourage Suppliers to propose "innovative WASH interventions"?	Not likely; PBR risks and tight timeline discouraged innovation and encouraged Suppliers to adopt (mostly) familiar, tested approaches. However, programme management was flexible and responsive to changes on the ground, with many examples of results-oriented problem-solving.	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews **Secondary** • DFID documentation	Evidence of clear intent of the programme to encourage 'innovative' WASH interventions, inclusive results and learning, based on DFID documentation. Assessment of the extent to which each was achieved by the end of the Output Phase, based on triangulation of perceptions of DFID, Suppliers, and unsuccessful bidders.	4.8.1	
DEQ 1.7: How likely was it that the programme design would encourage inclusive outputs and outcomes?	Unclear at this stage; Suppliers have targeted under-served locations but data unavailable on extent to which results so far have been inclusive and benefitted the lowest wealth quintile. DFID did not specify a targeting process or enforce the need to report on disaggregated results.			4.8.3	
DEQ 1.8: How appropriate was the WASH Results Programme's design for achieving the programme 'learning objectives'?	Somewhat appropriate; there has been some learning at global level and within Supplier organisations, but tight timeline left insufficient space for this, and no explicit lesson-sharing mechanism was built in at the start.			6.5	**HIGH** Minor risk of obtaining different opinions within DFID
DEQ 1.9. To what extent was the **design of each consortium sub-programme** appropriate for achieving DFID's key objectives?	Designs were appropriate insofar as they enabled delivery of the scale and type of results sought by DFID. However, equity, sustainability and government ownership were not built into payment criteria, and in at least some projects these aspects were marginalised during the Output Phase.	**Primary** • Lead Supplier interviews • Supplier country case studies • Country manager survey **Secondary** • Quarterly verification reports	Evidence of consistency between the objectives of DFID's design and the design of the Supplier programmes, based on triangulation of perceptions of DFID and Supplier staff. Supplier tender documents were also reviewed, but in many cases this did not give a full picture of the interventions implemented.	4.5	
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?	Each Supplier has achieved its total Output targets, but few sub-projects target entire districts/sub-districts or operate via the government framework. 'At scale' was not defined in the programme design.	**Primary** • Supplier country case studies **Secondary** • Quarterly verification reports	Quantitative assessment of % of output targets achieved in each country by 2016 (disaggregated by implementing partner if possible). Assessment of the 'scale' targeted by each partner, based on interviews with Supplier staff. Supplier tender documents were also reviewed, but in many cases this did not give enough detail on the scale of the intervention.	6.1.1	**HIGH**

		• Supplier monitoring data			
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)?	To be assessed in 2018.	**Primary** • Lead Supplier interviews • Supplier country case studies			
DEQ 2.3: To what extent have services continued to function and have behaviours continued to be used since their initial implementation (sustainable outcomes)?	To be assessed in 2018.	• Country manager survey • Local partner interviews • Community-level FGDs and KIIs • Pakistan quant-qual research study **Secondary** • Quarterly verification reports • Supplier monitoring data • Supplier baseline/endline data	Quantitative assessment of % of outputs sustained until endline (disaggregated by implementing partner if possible). Quantitative assessment of % of countries with inclusive outcomes at endline (disaggregated by implementing partner if possible). Assessment against the prerequisites for sustainable and equitable WASH outcomes (as set out in Design Document).		**MEDIUM** May not obtain access to Supplier monitoring data and baseline/ endline data Challenging to assess validity of evidence across highly varied country experiences
DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?	A critical challenge for all Suppliers was the tight MDGs deadline; nevertheless, targets were met. Complexity of consortium membership was also key. To be further explored in 2018.	**Primary** • Lead Supplier interviews • LV interviews • Supplier country case studies	Review of the risk factors cited in the quarterly Verification Reports as affecting implementation in under- and over-performing countries. Assessment of the main factors affecting implementation across under- and over-performing countries, based on triangulation of perceptions of DFID, Suppliers, and Verifiers.	6.1	**MEDIUM** Contingent on interviewees having a detailed knowledge of driving factors across different countries. Contingent on observing different levels of achievements across countries

DEQ 2.5: Under which circumstances did the **PBR framework** help/hinder the achievement of intended outputs and outcomes?	PBR risks sharpened minds on results and each Supplier over-delivered outputs. However, the need for cash flow also led Suppliers to include payment for inputs in their results frameworks, which added considerably to the reporting and verification burden.	• Country manager survey • National KII • Local partner interviews • Community-level FGDs and KIIs	Review of the PBR effects cited in the quarterly Verification Reports. Evidence of examples of how PBR affected the achievement and quality of outputs (explanation of causal links / external drivers based on staff experience with previous non-PBR programmes), based on triangulation of views across Supplier staff in country and at HQ.	6.1	**MEDIUM** Contingent on interviewees being able to compare experiences to a relevant non-PBR programme
DEQ 2.6: Under which circumstances did the **PBR framework** affect the quality of programme implementation (positive or negative)?	The quality of processes and outputs was generally good. But in a few isolated cases, the MDG deadline and associated payment risks led implementing agencies to adopt a 'light' version of their operational model or waive established procurement rules.			6.1, 6.2	
DEQ 2.7: Under which circumstances did Suppliers implement innovative approaches and focus on learning?	Little evidence of Suppliers adopting innovative operational approaches, though management was flexible and responsive to local circumstances. On learning see 1.8 above.	**Secondary** • Quarterly verification reports	Evidence of examples of innovation (compared to staff experience with previous non-PBR programmes), based on triangulation of views across Supplier staff in country and at HQ.	6.1.4	**HIGH**

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?	Many aspects of the global programme design were unclear at the bidding stage. This resulted in a long procurement process and inception period, which – in addition to the long design phase – reduced the time available for implementation.	**Primary** • DFID interviews • Lead Supplier interviews • Unsuccessful bidder interviews	Evidence of transaction costs, based on triangulation of perceptions of DFID, Suppliers, and unsuccessful bidders. Evidence of consequences resulting from transaction costs, based on triangulation of views across Supplier staff in country and at HQ.	2.2.3	**HIGH** Minor risk of not achieving a balanced view of the tendering process between bidders and DFID Minor risk of not interviewing key stakeholders who have left (institutional memory)
DEQ 3.2: Under which circumstances did the **PBR modality** affect the cost-effectiveness of individual sub-programmes?	PBR allowed flexible use of budgets but also created cash flow challenges, as Suppliers had to pre-finance activities. Reporting and verification burden was considerable, creating significant additional costs. Data unavailable on whether results were delivered with greater efficiency or VFM than would be the case with other financing modalities.	**Primary** • Lead Supplier interviews • Supplier country case studies • Country manager survey	Evidence of examples of how PBR affected costs and how PBR strengthened M&E arrangements (compared to staff experience with previous non-PBR programmes), based on triangulation of views across Supplier staff in country and at HQ. Evidence of examples of how strengthened M&E arrangements led to improved implementation. Quantitative assessment of the staff time associated with verification burden in country case studies (not feasible in 2016; will be done in 2018).	6.4	
DEQ 3.3: Under which circumstances did the **PBR modality** strengthen the programme monitoring and management arrangements of individual sub-programmes?	Suppliers had to strengthen their monitoring and reporting systems to meet verification requirements. This provided DFID with greater confidence regarding the results reported. However, it is still uncertain at this stage whether the onerous monitoring and reporting burden added value to the quality of programme implementation.			6.2.2	**HIGH**

DEQ 3.4: Under which circumstances did **key programme features** affect cost-effectiveness?	To be assessed in 2018.		VFM assessment comparing the cost and price per beneficiaries and isolating key drivers of cost-effectiveness.		
DEQ 3.4: Under what circumstances did consortium complexity affect the efficiency of the **programme management** arrangements of individual sub-programmes?	Programme management burden increased with the number of partners and range of programme activities.	**Primary** • DFID interviews • Lead Supplier interviews • Supplier country case studies	Evidence of how consortium management arrangements affected implementation, based on triangulation of views across Supplier staff in country and at HQ, and DFID.	6.3	**HIGH** Minor risk of not achieving a balanced view from programme implementers

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**?	Uncertain; will consider further in 2018. However, it is notable that DFID did not require the achievement of ODF status or the concentration of water, sanitation and hygiene interventions in the same locations.	**Primary** • Lead Supplier interviews • LV interviews • Supplier country case studies	Health Assessment against the prerequisites for WASH health impacts (as set out in Design Document). Non-health Evidence of non-health impacts (explanation of causal links / external drivers).	**7.4.3**	**MEDIUM** Challenging to assess validity of evidence across highly varied country experiences
DEQ 4.2: Under which circumstances did the WASH Results Programme activities have any **unintended/ unplanned positive or negative impacts**?	Uncertain; will consider further in 2018. Considerable staff turnover in some projects and anecdotal evidence that in some cases this was the result (at least in part) of stress related to tight deadlines and onerous reporting requirements. Staff turnover has negative implications for programme continuity and institutional memory.	• Country manager survey • National KII • Local partner interviews • Community-level FGDs and KIIs **Secondary** • Quarterly verification reports • WASH sector review for case studies	Review of any unintended effects cited in the quarterly Verification Reports. Evidence of factors which caused unintended impacts (explanation of causal links / external drivers).	6.1.8	**MEDIUM** Contribution of the WASH Results Programme to unintended impacts is harder to assess at community-level Challenging to assess validity of evidence across highly varied country experiences

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?	Risks are evident: groundwork for sustainability marginalised in some projects due to tight output deadline; having separate Output / Outcome phases may also have encouraged this inadvertently. To be further explored in 2018.	**Primary** • Lead Supplier interviews • LV interviews • Supplier country case studies • Country manager survey • Local partner interviews • Community-level FGDs and KIIs **Secondary** • Quarterly verification reports	Review of the risks cited in the sustainability RAG ratings in each Verification Report. Assessment against the prerequisites for sustainable WASH outcomes (as set out in Design Document), based on triangulation of examples given by DFID, Suppliers, national sector stakeholders and beneficiaries.	**7.4.2**	**MEDIUM** Challenging to assess validity of evidence across highly varied country experiences
DEQ 5.2: Under which circumstances has the **PBR modality** affected the likelihood of long-term sustainability of the outcomes and impacts?	To be assessed in 2018.	*As above +* • National KII	Evidence of programme features affecting the likelihood of sustainability (explanation of causal links / external drivers), based on triangulation of perceptions of DFID, Suppliers, national sector stakeholders and the MV team.		**MEDIUM** Challenging to assess validity of evidence across highly varied country experiences
DEQ 5.3: Under which circumstances have **other programme features** affected the likelihood of the long-term sustainability of the outcomes and impacts?	To be assessed in 2018.				
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**?	To be assessed in 2018.	*As above +* • Interview with Itad L&D team	Evidence of learning; evidence of changes to policy; and evidence of WASH Results Programme contribution to better evidence-based policy (explanation of causal links / external drivers) – based on triangulation of perceptions of DFID, Suppliers, and the MV team.		**MEDIUM** Challenging to assess validity of evidence across highly varied country experiences

C.2   List of people interviewed

Table 17:   People interviewed during Pakistan case study

Name	Position	Organisation
Sajid Ayyub	Acting CCU Director	Plan International
Zia Ullah	Senior CCU Finance Officer	Plan International
Rashid Khan	CEO	Associates in Development (AiD)
Babur Ghani	Research Assistant	AiD
Sajil Liaquat	Health Adviser	DFID
Kamran Naeem	WASH Specialist	UNICEF
Shabahat Ambreen	WASH Specialist for Punjab	UNICEF
M. Irfan Saeed Alrai	WASH Specialist for Sindh	UNICEF
Muhammad Dawud Mufti	Project Manager – SAWRP	Plan International
Azam Kiani	BCC Coordinator – SAWRP	Plan International
Anum Sarwar	Knowledge and Learning Coordinator	Plan International
Iftikhar Raja	Project Manager – SAWRP Bahawalpur	Plan International
Alyia Sundus	CLTS Trainer – SAWRP Bahawalpur	Plan International
Saima	MER Officer – SAWRP Bahawalpur	Plan International
Hasan	SanMart Officer – SAWRP Bahawalpur	Plan International
Ayesha	SanMart Officer – PATS Bahawalpur	Plan International
Siddiq Ahmed Khan	Country Representative	WaterAid
Sohail Nazir	Manager Programme Implementation	WaterAid
Imran Ali Shah	Project Manager – SAWRP	WaterAid
Faisal Gorchani	BCC manager – SAWRP	WaterAid
Sabir Hussain	M&E Coordinator - SAWRP	WaterAid
Sarfraz Ali Memon	Programme Coordinator Sindh	WaterAid
Ammar Mohsin	Business manager – Lifebuoy	UNILEVER
Waqas Hameed Bajwa	Deputy Director Marketing and Partnership	Idara-e-Taleem-o-Aagahi (ITA)
Ejaz Sherazi	District Project Coordinator	ITA
Habib Ahmed	CEO	Lodhran Pilot Project (LPP)
Akhtar Hussain	Project manager – SAWRP	LPP
Tasneem Aslam	MER Officer – SAWRP	LPP
Waseem Akhtar	District Manager SAWRP Lodhran	LPP
Mohd Sabir	SAWRP Project Engineer	LPP
Abdul Salam	WASH Coordinator (National)	NRSP
Nazar Joyo	Project Manager	NRSP – Thatta
Ahmed Ali Soumro	Project Officer MER	NRSP – Thatta
Syed Hassan Ahmad	District Coordinator SAWRP/NRSP	NRSP – Bahawalpur
Muhammad Usman	MER Officer SAWRP	NRSP – Bahawalpur
Syed Kazim Kirmani	Manager Special Operations and Special WASH Programmes	Muslim Aid
Babar Ali	Project Manager	Muslim Aid – Rajanpur
M. Bilal	MER Officer	Muslim Aid – Rajanpur
Najjeeb Aslam	Director of CD and Training	LGCD Province Punjab
Faheem Akhter Junejo	Director General (M&E)	LGCD Province Sindh
Rehana Ramzan	Community Based Motivator – PHED	Bahawalpur District Gov.

Naushee Kanwal	Community Based Motivator – PHED	Bahawalpur District Gov.
Mohd Khursheed	LGCD	Bahawalpur District Gov.
Mohd Irfan	Research Officer	Lodhran PHED district Govt.
K.B. Behrani	District Officer for Social Welfare and Secretary of WASH Forum	Thatta District Government
Dr Ghulam Sarwar	Medical Superintendent RHC – Gharo	Thatta District Government
Anwar Kalmati	Supervisor Education Department	Thatta District Government
Jan Mohd	Secretary Union Council Haj Gijano	Thatta District Government
Hidayafullah	Secretary Union Council Haj Gijano	Thatta District Government
Shahid Aziz	Administrator	Kute Kids School, Rawalpindi
Aasma Furqan	Teacher (So5 trained)	Kute Kids School, Rawalpindi

Table 18: **People interviewed during Uganda case study**

Name	Given name	Title / affiliation
SNV Uganda, Ethiopia and PMU Nairobi		
De Regt	Jeanette	Country Director, SNV Uganda
Mutta	Anne	SSH4A Programme Manager, SNV PMU, Nairobi
Nabasirye	Lillian	SSH4A Programme Manager Uganda
Getachew	Tessema	SSH4A Programme Manager, Ethiopia
Kemigisha	Cathy	Finance Officer, SNV Uganda
Lakwo	Dennis	Programme Adviser West Nile, SNV Uganda
Kakooza	Henry	Programme Adviser Rwenzori, SNV Uganda
Nakku	Stella	Field Operations Manager, SNV Uganda
Otim	Robert	Programme Adviser Rwenzori, SNV Uganda
Uganda Sanitation Fund		
Mukama	David	Programme Manager Uganda Sanitation Fund
Kitimbo	Jimmy	USF Field Officer, West Nile
Country Verifiers		
Katuramu	Patric	Country Verifier Uganda, Itad
Muktar	Abduke	Country Verifier Ethiopia, Itad
DFID		
Barnett	Ed	Education Adviser, DFID Kampala
Government of Uganda		
Kabirizi	Aaron	Director, Ministry of Water and Environment
Kyomuhangi	Julian	Assistant Commissioner, EHD, Ministry of Health
District Local Government		
Amandu	Alfred	District Health Inspector Zombo District, West Nile
Bagonza	Stephen	Acting District Health Inspector, Kasese
Mutooro	Julius	Health Inspector / Biostatistician, Kasese
Captain Mwesige	James	RDC, Kasese
Sakwa	David	Senior Health Inspector, Kibaale District
Kamara	James	Assistant CAO, Kyegegwa District
Mugweri	Thomas	Assistant District Water Officer Sanitation, Kyegegwa District
Baseka	Yusuf	District Health Officer, Kasese
Sub-county Officials		

Bwambale	Aprunale	LCIII Bugoye Sub-county, Kasese
Murongo	Esau	Sub-county Chief, Bugoye, Kasese
Senyonga	Vincent	Senior Assistant Secretary Mataale Sub-county, Kibaale District
Kayondo	Geoffrey	Rwentuha Sub-county Chief, Kyegegwa District

Table 19:   SWIFT partners interviewed during Kenya case study

Organisation	Persons met	Position and contact
Concern Worldwide	Winniefed Mwenii	FIM- WASH, SWIFT contact person +254725530336
Sanergy	Louise Couder	Business Development Associate +254705428909 Louise.couder@saner.gy
	Ruthie Rosenberg	Ruthie.rosenberg@saner.gy
	Polycarp	School WASH Coordinator
WSUP	Eden Mati	Project Manager +254718393968 emati@wsup.com
	Silvia Ndirangu	Project officer +254718392950 sndirangu@wsup.com
	Getrude Salano	MEAL Officer +254718392302 gsalano@wsup.com
Practical Action	Eng. Peter Murigi	Urban WASH Specialist +254722458516 Peter.murigi@practicalaction.or.ke
	Jane Njomo	Snr Project Officer – FADRR +254707997377 Jane.njomo@practicalaction.or.ke
	Francis Muchiri	Communication officer +254721719676 Francis.muciri@practicalaction.or.ke
	Caroline Gathu	Project assistant – FADRR +254722775049 Caroline.gathu@practicalaction.or.ke
BBC Media Action	David Njuguna	Broadcast mentor +254720803587 david.njuguna@bbcmediaaction.org
	Diana Njeru	Project Manager +254722162021 Diana.njeru@bbcmediaaction.org
Oxfam	Simeon Ogamba	SWIFT Project Coordinator/WASH Engineer
	James Ndirangu	Grant Accountant – Kenya programme
	Sepharina Owino Onyango	Public Health Engineer, Lodwar
	Ruth Tino	Programmes coordinator, Lodwar

Table 20:   KIIs during Kenya case study

Name	Address	Organisation

Fidelis Nene	Community Development Assistant (CDA)	Nairobi Water Company, Dandora
Jackline Wangeci	Consultant for promotion of connections	WSUP, Dandora
Livingstone Gichamu	Village Chairman	Dandora Water project
Ignitious Mirigu	Promoter of metered connections	WSUP/Dandora Water Project
David Koskei	Team leader, drilling and equipping +254725429337	Turkana county government
Paul Ekron	Commercial and Admin Manager +254729846080	Lodwar Water Supply Company (LOWASCO)
Francis Adome	Technical manager +254710241075 fradome601@gmail.com	(LOWASCO
Rael K. Akoru	Assistant chief public health officer and school health coordinator +254705275586, 0711897920 raelkukule@yahoo.com	Turkana County Government, Lodwar
James Kooya	Deputy public health officer +254720402516 kooyaj@yahoo.com	Turkana County Government, Kakuma
Francis Ekai Paule	Community Health Volunteer +254705582985	Ngimunyanakirionok / Ngikwakais village, Kakuma
Vincent Loochok	Community Health Volunteer +254705098676	Ngimunyanakirionok / Ngikwakais village, Kakuma
Lopeyok Edapal	Community Health Volunteer +254705131720	Ngimunyanakirionok / Ngikwakais village, Kakuma
George Wasonga	Sub-county Water Officer +254722976845	Turkana County Government, Kakuma
Callinton Ekai Akerengimop	Chairman +254706171368 kawasepro@yahoo.com	Kakuma Water Services project (KAWASEPRO)
Pius Nzuki Kitonyi	Sanitation technician +254717493812 pnzuki@oxfam.org.uk	Oxfam, Lodwar
Audrey Andwati	Public health promoter +254725969039 aandwati@oxfam.org.uk	Oxfam, Lodwar
Masinde	Radio presenter	Local radio in Lodwar
Jane Njomo	Snr Project Officer – FADRR +254707997377 Jane.njomo@practicalaction.or.ke	Practical Action Based in Nairobi but serving Turkana county
Benedict Ekeru	Project Assistant / sanitation technician +254725059735 benedict.ekeru@practicalaction.or.ke	Practical Action, Lodwar
Joseph Ekiru Lotokos	Chairman and water operator +254728754312 josephlotokos@gmail.com	Nayanae-Atiir Water Users Committee, Lokichar (supported by Practical Action)
Michael Ayen	Vice chairman +254716732738	Nayanae-Atiir Water Users Committee, Lokichar (supported by Practical Action)

Regina Matuka	Treasurer	Nayanae-Atiir Water Users Committee, Lokichar (supported by Practical Action)
William Betel	Chairman +2547145698	Nangorchoto Water Users Association, Lokichar (supported by Practical Action)
Winniefred Mwenii	FIM- WASH, SWIFT contact person +254725530336	Concern Worldwide based in Nairobi and serving Marsabit Office
Loice Wairimu	Accounts Officer	Concern Worldwide based in Nairobi and serving Marsabit Office
Jones Arada	Procurement Officer	Concern Worldwide based in Nairobi and serving Marsabit Office
James Moturi	WASH Engineer +254724671580	Concern Worldwide Marsabit
Gothu Guyo Abudho	Sanitation and Hygiene Officer +245721855126 gothu.guyo@concern.net	Concern Worldwide Marsabit
Josephine K. Mulaa	Senior Manager ASAL FIM WASH programme	Concern Worldwide Marsabit
Amina Isako	WASH Manager +254727487056	Pastoralist Community Initiative and Development Assistance (PACIDA) Local partner for Concern Worldwide in Marsabit
Galgallo Tanda	Water officer +254726080157	PACIDA Local partner for Concern Worldwide in Marsabit
Denge Boru	Hygiene Promoter	PACIDA Local partner for Concern Worldwide in Marsabit
Henry Halaki	Project Coordinator	Anglican Development Services (ADS) Local partner for Concern Worldwide in Marsabit
Wario Mamo	Chairman of committee	Turbi Rural water Supply Project in Lagh Warabesa. One of the supported communities
Sheikh Mohamed Wario	Committee member	Turbi Rural water Supply Project in Lagh Warabesa
Benson Thuku	Area Coordinator Marsabit +254721344053 Benson.thuku@concern.net	Concern Worldwide Marsabit
Baraka G. Katelo	Public Health Officer +254725746280 barakogic@gmail.com	Marsabit County government
Innocent Sifuna	Public Health Officer +254 717257979 innocentmunyefu@gmail.com	Turkana County Government
Chamia Mutua	Verifier (WASH M&E)	OPM
Diana Ogero	WASH Engineer +254715148182	Practical Action, Lodwar
Kontoma Thomas	Project manager/WASH officer +254724328180	Oxfam Wajir
Mohamed Salat	Project officer/public health promoter +245722404358	Arid Lands Development Focus (ALDEF) a local partner with Oxfam in Wajir

Adbiwahid Ahmed	Programme Officer/ Public Health promoter +254722761046	Wajir South Development Association (WASDA) a local partner with Oxfam in Wajir
Mr. Roble	+254710968680	Wajir Water and Sanitation Company (Wajwasco) officer
Mr John Kigotho Chege	Community Development Officer +254710115268	Nairobi Water Company

Table 21:   Interviews held at global level

The interviews held with unsuccessful bidders in November 2014 are not cited below, for reasons of confidentiality.

Organisation	Name	Role
DFID	Leonard Tedd	Senior Responsible Owner (SRO)
	Guy Howard	Head of WASH Team
	Jim McIntyre	Procurement
ITAD	Kathi Welle	Deputy-lead of MV Team
	Rachel Norman	Deputy-lead of MV Team (maternity cover)
	Andrew Robinson	LV for SSH4A
IWEL	Don Brown	Lead of MV Team
	Alison Barrett	LV for SAWRP
	Joe Gomme	LV for SWIFT
Plan	Marc Faux	Lead for SAWRP
SNV	Antoinette Kome	Lead for SSH4A (2014)
	Anne Mutta	Lead for SSH4A (2015-2018)
Oxfam	Francesco Rigamonti	Lead for SWIFT (2014-2016)
	Joanna Trevor	Lead for SWIFT (2016-2018) DRC country manager (2014-2016)

Table 22:   Webinars / telephone interviews held with country managers

Organisation	Country manager position	Name
SNV	Ethiopia	Andualem Anteneh
	Ghana	Jesse Danku
	Kenya	Fanuel Nyaboro
	Mozambique	Temesgen Befekadu
	South Sudan	Moses Kamanga
	Tanzania	Jackson Wandera
	Zambia	Solomon Mbewe
	Uganda	Lilian Nabasirye
Oxfam	Brian McSorley	Kenya country manager (2014–2016)
	Joanna Trevor	DRC country manager (2014–2016)

Annex D   Country case study reports

D.1   Pakistan report (SAWRP)

D.2   Kenya report (SWIFT)

D.3   Uganda report (SSH4A)

D.4   Supplementary study of PBR arrangements between SNV Uganda and implementing partners